Federal Reserve Bank of Cleveland Working Paper Series

Constructing Fan Charts from the Ragged Edge of SPF Forecasts

Todd E. Clark, Gergely Ganics, and Elmar Mertens

Working Paper No. 22-36

November 2022

Suggested citation: Clark, Todd E., Gergely Ganics, and Elmar Mertens. 2022. "Constructing Fan Charts from the Ragged Edge of SPF Forecasts." Working Paper No. 22-36. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202236.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Constructing Fan Charts from the Ragged Edge of SPF Forecasts*

Todd E. Clark,[1] Gergely Ganics,[2] and Elmar Mertens[3]

[1] *Federal Reserve Bank of Cleveland,* [2] *Banco de España and John von Neumann University,* [3] *Deutsche Bundesbank.*

November 20, 2022

Abstract

We develop a model that permits the estimation of a term structure of both expectations and forecast uncertainty for application to professional forecasts such as the Survey of Professional Forecasters (SPF). Our approach exactly replicates a given data set of predictions from the SPF (or a similar forecast source) without measurement error. Our model captures fixed-horizon and fixed-event forecasts, and can accommodate changes in the maximal forecast horizon available from the SPF. The model casts a decomposition of multi-period forecast errors into a sequence of forecast updates that may be partially unobserved, resulting in a multivariate unobserved components model. In our empirical analysis, we provide quarterly term structures of expectations and uncertainty bands. Our preferred specification features stochastic volatility in forecast updates, which improves forecast performance and yields model estimates of forecast uncertainty that vary over time. We conclude by constructing SPF-based fan charts for calendar-year forecasts like those published by the Federal Reserve.

Keywords: Term structure of expectations, uncertainty, survey forecasts, fan charts
JEL classification codes: E37, C53

*Parts of this paper were earlier circulated under the title "Constructing the Term Structure of Uncertainty from the Ragged Edge of SPF Forecasts" and were completed while G. Ganics was with the Central Bank of Hungary. We gratefully acknowledge helpful suggestions and comments received from Cem Çakmakı, Refet Gurkaynak, Mike McCracken, James Mitchell, and workshop or conference participants at the NBER Summer Institute 2022, 2022 IAAE conference, 2022 Dolomiti Macro Meeting, 2022 NBER-NSF SBIES conference, 2022 NBER-NSF Time Series conference, 2022 Conference on Real-Time Data Analysis, Methods, and Applications, Heidelberg University, and Deutsche Bundesbank. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland, the Federal Reserve System, the Banco de España, the Deutsche Bundesbank, or the Eurosystem. Replication files are available at `https://github.com/elmarmertens/ClarkGanicsMertensSPFfancharts`.

1 Introduction

The macroeconomic projections of professional forecasters are widely used in both economic policymaking and forecasting research. Such forecasts with long histories include the (US) Survey of Professional Forecasters (SPF), Blue Chip Consensus, Consensus Economics, and IHS Markit (formerly Macroeconomic Advisers), as well as Federal Reserve forecasts published in the Tealbook or Greenbook and the Federal Open Market Committee's (FOMC) Summary of Economic Projections. These forecasts have a wealth of useful information, but with some unevenness in availability across forecast horizons. For example, the SPF includes both (1) fixed-horizon quarterly point forecasts, at shorter horizons, and (2) fixed-event annual forecasts, covering shorter and longer horizons.[1] We characterize this available information as the "ragged edge of SPF forecasts."

A number of studies have developed models that use the mixtures of forecast horizons available in professional forecasts to build out a more complete term structure (across horizons) of point forecasts. For example, Aruoba (2020) and Kozicki and Tinsley (2012) develop models to extend point forecasts to obtain a more complete term structure of inflation forecasts. Focusing on capturing time variation in long-run forecasts, Crump, et al. (2021) develop a multivariate unobserved components model of trend and cycle to fit a range of macroeconomic survey forecasts to estimate a term structure of point forecasts in growth, inflation, and a short-term interest rate. In a similar vein, Hepenstrick and Blunier (2022) interpolate a term structure of complete quarterly forecasts from available quarterly and annual forecasts using a state-space representation that includes a simple time series process for quarterly growth and growth forecasts and a measurement equation that relates the forecasts of interest to the available measurements. Some other work (e.g., Dovern, Fritsche, and Slacalek (2012)) has taken a simpler interpolation approach to translate fixed-event point forecasts to fixed-horizon predictions, and Ganics, Rossi, and Sekhposyan (2021) develop an approach for translating fixed-event density forecasts to fixed-horizon quarterly forecasts.

We develop a model that permits not only the estimation of a term structure of expectations, but also a term structure of forecast uncertainty. Forecast uncertainty is widely recognized to be

[1] The available sample of forecasts from the SPF is also longer than those of most other professional forecasts.

important for monetary policy decisions, and many central banks (e.g., the Bank of England and the FOMC) publish estimates of their forecast uncertainty as an integral part of their policy communications. Clark, McCracken, and Mertens (2020) (hereafter, CMM) develop a model that uses realized errors in SPF point forecasts to estimate forecast uncertainty, pooling information embedded in forecast errors for different forecast horizons. Echoing results from prior research on model-based forecasts — typically based on vector autoregressions (VARs) and dynamic stochastic general equilibrium (DSGE) models — CMM find that model fit and forecast performance are improved by allowing for time-varying conditional variances. Specifically, CMM employ a multivariate stochastic volatility specification that improves the accuracy of uncertainty measures for survey forecasts compared to simpler approaches for tracking variances.[2] While CMM can be seen as yielding a term structure of forecast uncertainty, the maximal horizon of their application is limited to the four-quarter fixed-horizon forecasts of the SPF.

In this paper, to obtain a longer and more complete term structure of forecasts and their uncertainty, we develop new models that extend CMM in three directions. First, we generalize the model to include not only fixed-horizon forecasts but also fixed-event forecasts, including at horizons beyond the fixed-horizon maximum of four quarters ahead. Our methods allow us to exactly replicate a given data set of predictions (from the SPF or other judgmental forecast sources) without measurement error. Second, we also generalize the model to accommodate the ragged edge of SPF forecasts due to fixed-event horizons varying systematically within each year of the quarterly publication of the survey and due to the survey adding more years of fixed-event horizons some years ago. Together, these two model innovations permit us to use the ragged edge of SPF forecasts to obtain complete quarterly forecast fan charts at each forecast origin, extending to a horizon of 15 quarters. Third, we develop an extended version of the baseline model that allows for bias in observed SPF forecasts, resulting in persistent forecast errors. Time-varying bias in SPF

[2]Examples of VAR models with stochastic volatility include Clark (2011), Clark and Ravazzolo (2015), and D'Agostino, Gambetti, and Giannone (2013). Reifschneider and Tulip (2019) provide simple evidence of changes in the sizes of forecast errors associated with projections from the Federal Reserve and other sources, including the SPF and Blue Chip Consensus.

forecasts has been highlighted in studies including Bianchi, Ludvigson, and Ma (2022), Coibion and Gorodnichenko (2015), and Farmer, Nakamura, and Steinsson (2022). Our extended model treats SPF predictions as unbiased martingales only in its prior, while allowing for generic VAR dynamics of SPF forecasts in its posterior. The tightness of the prior can be set as tight or as loose — with the former pushing the model toward consistency with a martingales assumption and the latter permitting persistence in the SPF forecast errors — as the researcher prefers.

In our empirical analysis, we examine SPF forecasts of GDP growth, unemployment, inflation, and the 3-month Treasury bill rate. We first show that our model yields quarterly real-time forecasts that perfectly match and interpolate through the annual fixed-event point forecasts. We next show that the model's estimates of uncertainty vary over time and generally rise with the forecast horizon. Our third set of results addresses historical forecast accuracy. These results establish benefits — particularly for density forecasts — to including stochastic volatility in the model. They also show that the historical accuracy of point and density forecasts from the version of our model extended to allow for bias in observed SPF forecasts is generally very similar to the accuracy of our baseline model. So while on the one hand there is little benefit to generalizing our baseline model to allow for bias in SPF point forecasts, on the other there is little cost in forecast accuracy to doing so. This result is consistent with the existing literature that finds deviations from SPF forecast efficiency to be episodic and hard to consistently exploit in real time.[3]

We conclude the paper with a practical application using our estimated term structure of forecasts: constructing fan charts with calendar-year (four-quarter growth rates or fourth quarter levels) forecasts like those published by the FOMC. We show that, in recent years, our model applied to SPF forecasts yields estimates of uncertainty around GDP growth and unemployment forecasts that are lower than those implied by the historical forecast root-mean-squared errors (RMSEs) that underlie the FOMC's fan charts. Our approach could be used in real time with each new release of the SPF to produce updated fan charts of point forecasts and estimates of forecast uncertainty.

[3]See, for example, Croushore (2010), or more recently, Bianchi, Ludvigson, and Ma (2022), Foerster and Matthes (2022), Hajdini and Kurmann (2022), and Mertens and Nason (2020).

The paper proceeds as follows. Section 2 provides an overview of the related literature on survey forecasts not covered above. Section 3 describes the SPF forecasts and data used in the evaluation. Section 4 presents our model and briefly describes estimation. Section 5 provides results. Section 6 concludes. Additional results are provided in a separate supplementary online appendix.

2 Related Literature

A long literature — more than can be covered here — has examined whether professional forecasts display properties consistent with optimal (typically, under quadratic loss) forecasts and rational expectations. In one example, Patton and Timmermann (2012) develop new rationality tests based on rationality restrictions taking the form of bounds on second moments of the data across forecast horizons and apply them to forecasts from the Federal Reserve's Greenbook. Coibion and Gorodnichenko (2015) develop a new approach to testing rational expectations that permits quantifying departures from full rationality and the degree of information rigidities; their applications include inflation forecasts from the SPF. Focusing on inflation forecasts, Ang, Bekaert, and Wei (2007) find survey forecasts hard to beat by a battery of forecasting methods, and Croushore (2010) documents that deviations of the SPF (and the Livingston Survey) forecasts from rationality are typically short-lived and hard to exploit in real time. In this spirit, Mertens and Nason (2020) propose a new unobserved components model of inflation that distinguishes trend and inflation gap components and features a sticky information forecast mechanism; their estimates reveal that the stickiness of survey forecasts is not invariant to the time series process governing actual inflation. Using machine learning algorithms, Bianchi, Ludvigson, and Ma (2022) find evidence of time-varying bias in survey expectations and forecasts and conclude that artificial intelligence can be used to improve forecast accuracy. Regarding the predictive value of density forecasts collected by surveys, Clements (2018), as well as Glas and Hartmann (2022) and others, points to potential shortcomings, for example, due to rounding of answers by respondents.

Another long literature has sought to use professional forecasts to improve forecasts from time series models. In one example, Faust and Wright (2009) use professional forecasts as jumping-off points for models to improve the accuracy of forecasts from time series models. Wright (2013) shows that Bayesian VAR forecast accuracy can be improved by using long-run survey forecasts as priors on the long-run means of the model. Banbura, et al. (2021) and Krüger, Clark, and Ravazzolo (2017) improve forecasts from Bayesian VARs through entropic tilting toward survey-based nowcasts or forecasts. Frey and Mokinski (2016) instead add survey-based nowcasts as endogenous variables in Bayesian VARs, using priors so that the dynamics of the survey forecasts inform the parameter estimates of the dynamics of the actual data. In a similar vein, Doh and Smith (2022) develop priors that align a VAR's (a priori) forecasts with survey predictions.

Many other studies (in addition to those noted above) have used professional forecasts to examine the term structure of forecast uncertainty across horizons or time variation in uncertainty at given horizons. With data on fixed-event forecasts from Consensus Economics, Patton and Timmermann (2011) use an unobserved components model to examine the predictability of growth and inflation across different forecast horizons and measure average forecast uncertainty by mean-squared forecast errors. Clements and Galvão (2017) compare ex ante uncertainty estimates and ex post RMSEs from annual fixed-event SPF forecasts as well as corresponding measures from time series models of growth and inflation. To capture and assess time variation in macroeconomic uncertainty, Jo and Sekkel (2019) develop and estimate a factor stochastic volatility model using errors in SPF point forecasts of a small set of macroeconomic variables. Building on Adrian, Boyarchenko, and Giannone (2019), Adams, et al. (2021) find evidence of downside risks in the densities of observed SPF forecast errors based on quantile regressions that condition on financial conditions. In companion work (Clark, Ganics, and Mertens, 2022), we investigate the value of conditioning fan charts on information contained in the SPF's density forecasts, which come in the form of fixed-event probability bins. That approach also connects to the works of Bassetti, Casarin, and Del Negro (2022), Clements (2018), Clements and Galvão (2017), and Ganics, Rossi, and Sekhposyan (2021); for a recent survey, see Clements, Rich, and Tracy (2022).

3 Data

Because the availability of forecasts in the SPF informs aspects of our model, we detail the data in this section before taking up the model in Section 4. We examine quarterly and annual forecasts from the SPF for a basic set of major macroeconomic aggregates: real GDP growth (RGDP), the unemployment rate (UNRATE), inflation in the GDP price index (PGDP) and consumer price index (CPI), and the 3-month Treasury bill rate (TBILL, or T-bill).[4] (For simplicity, we use "GDP" and "GDP price index" to refer to output and price series, even though, in our real-time data, the measures are based on GNP and a fixed-weight deflator for some of the sample.) These variables are commonly included in research on the forecasting performance of models such as VARs or DSGE models. The various forecast sources analyzed by Reifschneider and Tulip (2019) cover a very similar set of variables. The SPF forecasts are widely studied, publicly available, and the longest available time series of forecasts for a range of variables. Alternatives such as the Blue Chip Consensus are not available publicly or for as long a sample.

[Table 1 about here.]

We obtained the SPF forecasts from the Federal Reserve Bank of Philadelphia's Real-Time Data Research Center. In all cases, we form the point forecasts using the average over all SPF responses. Reflecting the data available, our estimation samples start with 1969Q1 for GDP growth, unemployment, and GDP inflation and 1981Q4 for CPI inflation and the T-bill rate; the sample endpoint is 2022Q2 for all variables. At each forecast origin, the available fixed-horizon point forecasts typically span five quarters, from the current quarter through the next four quarters. Since

[4] The SPF defines the relevant unemployment and T-bill rate measures as the quarterly averages of monthly data. CPI inflation is computed as the percent change in the quarterly average level of the price index. For real GDP and its price index, the SPF solicits point forecasts in terms of levels, which we convert into growth rates. In order to map calendar-year predictions of growth rates pertaining to changes in yearly average levels into our model, we employ a log-linear approximation detailed in Section 4.3. Accordingly, we convert SPF point forecasts for real GDP and its price index into continuously compounded growth rates.

1981Q3, the SPF has included fixed-event point forecasts for the current and next calendar year.[5] In 2005Q3, the forecast horizon for CPI inflation was extended to include annual forecasts for one additional year, i.e., covering up to two calendar years ahead, and in 2009Q2, the forecast horizon was similarly extended for GDP growth, unemployment, and the T-bill rate to extend up to three years ahead. Table 1 lists the first available dates for SPF forecasts of different variables at different horizons.

Given the fixed-event nature of the SPF's calendar-year forecasts, the number of quarters until the end of the longest-horizon forecast varies over the course of a year. For example, the 2021Q4 SPF included fixed-event annual forecasts of growth and unemployment for the current year and the next three, so that the last annual forecast extends 12 quarters ahead (the annual forecast reported for 2024 includes 2024Q4, 12 quarters beyond the 2021Q4 forecast origin). In the 2022Q1 SPF, the last annual forecast for 2025 includes 2025Q4, 15 quarters beyond the forecast origin. These variations in the SPF's effective forecast horizon are also known as the "ragged edge." Our methods allow us to consistently construct fixed-horizon term structures that extend to 15 quarters out at every quarterly forecast origin, and thus extend beyond the ragged edge.[6]

The SPF uses certain conventions in its forecasts that we incorporate in the measurement specification of our model, as described further in Section 4. In terms of the SPF data we use, predictions for the unemployment rate and Treasury bill rate are expressed directly as quarterly and annual-average levels (depending on the forecast horizon), which can be mapped directly into our model. Similarly, for CPI inflation, the forecasts are reported as percentage changes in quarterly and Q4/Q4 index levels that are used as such in our model. However, for GDP growth and inflation

[5]As detailed in Section 4, our model omits data on current-year forecasts in our empirical analysis since their forecast horizon is subsumed by the available quarterly forecasts.

[6]As detailed further below, the 15-quarters horizon for the endpoint of our term structures applies to variables for which the SPF provides forecasts for three years ahead, like GDP, the unemployment rate, and the T-bill rate. For CPI inflation, the SPF provides only forecasts up to two years ahead, leading to a coverage of maximally 12 quarters; in the case of inflation in GDP prices, the maximal quarterly forecast horizon is 8 quarters as annual forecasts are only provided for the next year.

rates, the SPF solicits forecasts in levels, which we convert into growth rates. Specifically, depending on the forecast horizon, point forecasts pertain to quarterly or annual-average levels, which we convert to growth rates based on information included in the survey. For quarterly forecast targets, we use the lagged quarterly level as the basis. To obtain the next-year forecast of annual-average growth, we use the SPF‘s predictions for the current year as base values (and analogously for the forecasts two and three years ahead). For each forecast horizon, we calculate growth rates for GDP and the GDP price index as differences in log levels, as our model uses log-linearized expressions for quarterly and annual-average growth in these variables that are detailed in Section 4.

To estimate our model, we also need measures of the outcomes of the variables. In the case of GDP growth and GDP inflation, data can be substantially revised over time. Specifically, for GDP growth and GDP inflation, we obtain real-time measures for quarter $t-1$ data as these data were publicly available in quarter t from the quarterly files of real-time data in the Philadelphia Fed's Real-Time Data Set for Macroeconomists (RTDSM). For forecast evaluation, we measure the outcomes of GDP growth and GDP inflation with the RTDSM vintage published two quarters after the outcome date (that is, we use the quarterly vintage in $t+h+2$ to evaluate forecasts for $t+h$ made in t; this is the second estimate available in the RTDSM's vintages). Because revisions to quarterly data are relatively small for the unemployment rate and CPI inflation and non-existent for the T-bill rate, we simply use the historical time series available in the St. Louis Fed's FRED database to measure the outcomes and corresponding forecast errors for these variables.

4 Model and Estimation

We begin by specifying the general form of our baseline model and then proceed to explain its pieces and complexities in more detail. In broad terms, our model can be seen as a multivariate unobserved components model. We design the state-space specification to match arbitrary term structures of expectations, with application to the SPF in this paper. In our preferred specification, shocks to the model's state vector have stochastic volatility.

In all cases, we specify and estimate the model on a variable-by-variable basis (i.e., we estimate it separately for GDP growth, the unemployment rate, inflation, and the T-bill rate). The variable of interest in quarter t is denoted y_t, and forecasts for period $t+h$ from forecast origin t are denoted $y_{t+h|t}$. Let $\boldsymbol{Y_t}$ denote a (partially latent) state vector containing the lagged realized value and a term structure of quarterly fixed-horizon forecasts. Specifically, $\boldsymbol{Y_t}$ consists of the lagged realized value y_{t-1} (that is observed at t), the time t nowcast, and quarterly fixed-horizon forecasts extending from $h = 1$ up to the maximum quarterly horizon that can be covered in the historical SPF data. We denote this maximal horizon by H and obtain: $\boldsymbol{Y_t} = \left(y_{t-1}, y_{t|t}, y_{t+1|t}, \ldots, y_{t+H|t}\right)'$. Some of the elements of $\boldsymbol{Y_t}$ are unobserved. The available measures of SPF point forecasts — for quarterly horizons up to four steps ahead and for annual fixed events up to three years ahead — are collected in the measurement vector $\boldsymbol{Z_t}$.

4.1 An accounting identity for forecasts

We build our model from an accounting identity, also used by CMM, that decomposes h-step-ahead forecast errors into the sum of the $t+h$ nowcast error and preceding forecast updates:

$$y_{t+h} - y_{t+h|t} = e_{t+h} + \sum_{i=1}^{h} \mu_{t+h|t+i}, \tag{1}$$

with $e_{t+h} \equiv y_{t+h} - y_{t+h|t+h}$ and $\mu_{t+h|t+i} \equiv y_{t+h|t+i} - y_{t+h|t+i-1}$, so that e_{t+h} denotes the nowcast error at $t+h$, and $\mu_{t+h|t+i}$ measures the update in the forecast of y_{t+h} at time $t+i$.[7]

Our goal is to construct term structures of expectations and uncertainty for horizons $h = 0, 1, \ldots H$. We will also refer to the H-step-ahead forecast, $y_{t+H|t}$, as the long-run forecast. In the case of our application to the US SPF, the longest-available horizon H equals 15 quarters. In

[7]Some previous studies have also made use of expectational updates, for different purposes. For example, Patton and Timmermann (2012) write a short-horizon forecast as a sum of a long-horizon forecast and forecast revisions, and use it as the basis of an optimal revision regression to test forecast optimality (under quadratic loss and stationarity). Coibion and Gorodnichenko (2015) map out the implications of different theories of imperfect information for serial correlation in forecast updates.

an extension of the CMM framework, we also track the change in long-run forecasts from one quarter to the next, which is denoted as μ_t^*:

$$\mu_t^* \equiv y_{t+H|t} - y_{t+H-1|t-1}. \tag{2}$$

Without additional assumptions about the dynamic processes for e_{t+h}, $\mu_{t+h|t+i}$, or μ_t^*, equations (1) and (2) represent mere accounting identities (or definitional equations), with equation (1) also used by CMM. We collect $\mu_{t+h|t}$ (for all $0 \leq h \leq H$), μ_t^*, and e_{t-1} in a vector denoted $\boldsymbol{\eta}_t$.[8] For brevity, we refer to $\boldsymbol{\eta}_t$ as the vector of forecast updates.

Applied to forecasts $y_{t+h|t}$ for all $h \geq 0$, the accounting identities above lead to the following recursive representation of the state vector $\boldsymbol{Y_t}$:

$$\boldsymbol{Y_t} = \boldsymbol{F}\,\boldsymbol{Y_{t-1}} + \boldsymbol{\eta}_t, \tag{3}$$

$$\text{with} \quad \boldsymbol{Y_t} = \begin{bmatrix} y_{t-1} \\ y_{t|t} \\ y_{t+1|t} \\ \vdots \\ y_{t+H|t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \\ 0 & \dots & \dots & 0 & 1 \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}}_{\boldsymbol{F}} \boldsymbol{Y_{t-1}} + \underbrace{\begin{bmatrix} e_{t-1} \\ \mu_{t|t} \\ \mu_{t+1|t} \\ \vdots \\ \mu_t^* \end{bmatrix}}_{\boldsymbol{\eta}_t}.$$

The state vector $\boldsymbol{Y_t}$ contains the lagged value of y, whose first reading typically becomes available only in the current quarter, and quarterly SPF point forecasts from period t through $t+H$ (some or many unobserved). Equation (3) describes the evolution of forecasts and realized values, for a given sequence of forecast updates, $\boldsymbol{\eta}_t$. Critically, the transition matrix $\boldsymbol{F}$ is known and need not be estimated.[9] To close our model, we need to specify the dynamics of the forecast update vector $\boldsymbol{\eta}_t$, which we consider in the next subsection.

Although our model is written with a state vector $\boldsymbol{Y_t}$ containing SPF forecasts, we should em-

[8]The vector $\boldsymbol{\eta}_t$ includes the lagged nowcast error, since the realized value y_{t-1} is observed only at time t.

[9]All eigenvalues of the transition matrix $\boldsymbol{F}$ are 0, except for a single unit root.

phasize that, in our baseline model, we are not actually attributing a specific time series model to the evolution of SPF forecasts; we are taking the observed fixed-horizon and fixed-event forecasts as given and using a time series process to interpolate missing fixed-horizon forecasts out to H steps ahead. Assuming $\boldsymbol{\eta}_t$ is (mean-) stationary, our state equation embeds a common trend assumption, whereby survey forecasts and outcome variables share a single common trend, and survey forecast errors are (mean-) stationary. In addition, variations in the common trend component depend on the size of variations in μ_t^*. Our model can also represent the case when y_t is mean-stationary (and thus also $\boldsymbol{Y_t}$), by letting μ_t^* be 0 at all times; and for arbitrarily small variations in μ_t^*, the model provides a near-stationary representation of the data.

4.2 Transition equation in the martingale case

In our baseline specification, we embed an insight also underlying the original model of CMM: If point forecasts are optimal (under quadratic loss), then forecasts are martingales. As a consequence, forecast updates from one forecast origin to the next (with an unchanged target date) will be martingale differences. We thus assume that the expectational update $\mu_{t+h|t}$ forms a martingale difference sequence (MDS): $E_{t-1}\mu_{t+h|t} = 0$, and similarly for the nowcast error $E_{t-1}e_t = 0$.[10] In addition, we assume that changes in the long-run forecast are martingale differences, $E_{t-1}\mu_t^* = 0$. The latter assumption treats the long-run forecast, $y_{t+H|t}$, as the Beveridge-Nelson trend of y_t.[11] All told, our baseline model sees the vector $\boldsymbol{\eta}_t$ as a martingale difference sequence, $E_{t-1}\boldsymbol{\eta}_t = \mathbf{0}$. Throughout, we also assume that $\boldsymbol{\eta}_t$ is Gaussian (at least conditionally on yet-to-be defined time-varying parameters), which allows estimation based on standard sampling methods. In our baseline

[10] We use the expectations operator E_t to denote true expectations, under the average SPF respondent's information set, which is assumed to contain the econometrician's information set at time t. In the MDS case, observed SPF forecasts are assumed to be identical to true expectations, $y_{t+h|t} = E_t y_{t+h}$.

[11] As noted before, the relative variance of μ_t^* compared to the other innovations contained in $\boldsymbol{\eta}_t$ determines the strength of any non-stationary component in the data.

specification the state-space model has the following form:

$$\boldsymbol{Y_t} = \boldsymbol{F}\,\boldsymbol{Y_{t-1}} + \boldsymbol{\eta}_t\,, \qquad \boldsymbol{Z_t} = \boldsymbol{C}_t\,\boldsymbol{Y_t}\,, \qquad \boldsymbol{\eta}_t \sim N\left(\mathbf{0}, \mathrm{Var}_t\left(\boldsymbol{\eta}_t\right)\right)\,, \tag{4}$$

with the evolution of $\mathrm{Var}_t\left(\boldsymbol{\eta}_t\right)$ as a stochastic volatility (SV) process to be defined further below.

4.3 Measurement equation

To explain the measurement equation, we need additional notation for forecasts, covering the different types of annual forecasts (simple annual averages for the unemployment rate, CPI inflation, and the T-bill rate, and percent changes in annual averages for GDP growth and inflation in the GDP price index). To match forecasts of annual average levels and their growth rates, let

$$\bar{y}_t = 1/4 \cdot \sum_{j=0}^{3} y_{t-j}, \tag{5}$$

$$\hat{y}_t = 1/16 \cdot \left(y_t + 2 \cdot y_{t-1} + 3 \cdot y_{t-2} + 4 \cdot y_{t-3} + 3 \cdot y_{t-4} + 2 \cdot y_{t-5} + y_{t-6}\right), \tag{6}$$

so that when t corresponds to a date in Q4, $\bar{y}_t$ and $\hat{y}_t$ denote an annual observation.[12] We treat the calendar-year forecasts of the unemployment rate, CPI inflation, and T-bill rate as readings of $\bar{y}_{t+h|t}$, while annual forecasts for growth in real GDP and the GDP price index are read as data on $\hat{y}_{t+h|t}$. The formula for $\hat{y}_t$ represents a log-linear approximation to the growth rate in average levels of the years (or four-quarter intervals) ending at t and $t-4$.[13] Other examples relying on such approximations include Aruoba (2020), Mariano and Murasawa (2003), and Patton and Timmermann (2012). The variables $y_{t+h|t}$, $\bar{y}_{t+h|t}$, and $\hat{y}_{t+h|t}$ denote survey expectations collected at forecast origin t for forecast targets y_{t+h}, $\bar{y}_{t+h}$, and $\hat{y}_{t+h}$, respectively. At a given point in time,

[12] The use of a scale factor of 16 in the denominator of equation (6) reflects our definition of y_t as an annualized rate of change at quarterly frequency. If y_t was defined to measure a quarterly rate of change, the appropriate scale factor would be 4.

[13] In addition, we approximate the annual forecasts of percentage changes in Q4/Q4 levels of the CPI index as an arithmetic four-quarter average.

t, the survey data are assumed to provide observations of $y_{t+h|t}$, $\bar{y}_{t+h|t}$, and/or $\hat{y}_{t+h|t}$, for different (but separate) values of $h \geq 0$.[14] At different points in time, survey forecasts for different horizons h may be available. As noted above, the longest forecast horizon H covered by our model reflects the availability of annual forecasts. The quarterly fixed-horizon forecasts from the SPF extend only 4 quarters beyond the origin t (so this is the maximum horizon used by CMM). By making use of annual forecasts for up to three calendar years out, our term structures cover 16 quarters, ranging from the current quarter, $h = 0$, to $H = 15$.[15]

The measurement vector $\boldsymbol{Z_t}$ of the model contains readings from the SPF about forecasts for fixed horizons, $y_{t+h|t}$, or fixed events, $\bar{y}_{t+h|t}$ and $\hat{y}_{t+h|t}$, respectively, as well as a reading of the last realized value, y_{t-1}, that are available at a given point in time t for different horizons h. Since the SPF provides fixed-horizon forecasts for up to four quarters ahead, we disregard all current-year (fixed-event) forecasts. In addition, when the SPF is conducted in the fourth quarter of the current year, we also disregard the next-year forecast.[16] Otherwise, we include in $\boldsymbol{Z_t}$ all available readings of fixed-event forecasts for the next year and beyond. With this specification, the measurement equation takes the form $\boldsymbol{Z_t} = \boldsymbol{C}_t \, \boldsymbol{Y_t}$, where the elements of $\boldsymbol{C}_t$ are time-varying to reflect shifting data availability across surveys and over time (with the SPF adding additional years of annual forecasts as indicated above) but known.

Our approach differs in some notable ways from related work. Assuming rational expectations, one approach for fitting an arbitrary term structure of survey forecasts is to assume a time series model for y_t that generates forecasts $E_t y_{t+h}$ and to assume that observed survey forecasts are equal to model-implied forecasts plus a measurement error, $y_{t+h|t} = E_t y_{t+h} + \text{noise}_t$, and likewise for $\bar{y}_{t+h|t}$ and $\hat{y}_{t+h|t}$. This approach is employed by studies such as Aruoba (2020), Crump,

[14] Typically, a given survey source, like the (US) SPF, provides readings on $\bar{y}_{t+h|t}$ or $\hat{y}_{t+h|t}$, plus (possibly) $y_{t+h|t}$.

[15] In the case of inflation measured by CPI and GDP prices, SPF forecasts are elicited only up to two years and one year ahead, respectively, and we set $H = 11$ and $H = 7$ in those cases.

[16] We compared the SPF's current-year forecasts with their implied counterparts constructed from lagged data and the SPF's fixed-horizon forecasts and found any differences to be typically small, and similarly so for next-year forecasts made in the fourth quarter of a given year. While the sources of such differences may otherwise be of independent interest, they do not appear germane to our study.

et al. (2021), Grishchenko, Mouabbi, and Renne (2019), Kozicki and Tinsley (2012), and Mertens and Nason (2020). Potential drawbacks of this approach include the attribution of part of the observed term structure of survey expectations to measurement error and the imposition of a (typically low-order) time series model on the term structure of "true" expectations ($E_t y_{t+h}$).[17] For some applications, such an approach might provide a potentially beneficial form of shrinkage. In addition, the approach can be used to pool surveys from different sources to extract a common set of underlying expectations. However, the measurement error approach invariably comes at the cost of discarding part of the survey respondents' judgment and broader modeling. Instead, as noted above, in our baseline model we aim to perfectly fit model-implied forecasts to survey forecasts, $y_{t+h|t} = E_t y_{t+h}$, and our extended model sees differences between forecasts from the SPF and the model as bias (and thus predictable forecast errors, instead of measurement error).

4.4 Stochastic volatility

The stochastic volatility process of the innovation vector $\boldsymbol{\eta}_t$ distinguishes the last component of $\boldsymbol{\eta}_t$ that is the change in the long-run forecasts from one quarter to the next from the other components comprised of forecast updates at shorter horizons. We assume that the change in the long-run forecasts affects all forecasts at shorter horizons and has a variance that is constant over time. We assume a scalar volatility process that is common to all horizons except the longest, H. With these assumptions, the model decomposes forecast updates into long-run shifts and cyclical gaps:

$$\boldsymbol{\eta}_t = \begin{bmatrix} \tilde{\boldsymbol{\eta}}_t \\ 0 \end{bmatrix} + \mathbf{1} \cdot \mu_t^*, \qquad \mu_t^* \sim N(0, \sigma_*^2), \qquad \tilde{\boldsymbol{\eta}}_t \sim N\left(\mathbf{0}, \lambda_t \cdot \tilde{\boldsymbol{\Sigma}}\right), \tag{7}$$

$$\log \lambda_t = \delta \log \lambda_{t-1} + \nu_t, \qquad \nu_t \sim N(0, \sigma_\nu^2), \tag{8}$$

[17] Survey respondents (or the average survey respondent) may generate persistent forecast errors, consistent with information frictions. For simplicity, this case is ignored in our baseline specification. However, as detailed in Section 4.5, our approach can readily be extended to handle the case of persistent survey errors using a VAR specification of the $\boldsymbol{\eta}_t$ vector as in CMM.

where $\tilde{\boldsymbol{\eta}}_t$ is a vector of forecast-update gaps with $N_y - 1$ elements.[18] Carriero, Clark, and Marcellino (2016) develop a common volatility model for use with VARs in which volatility comovement is high. In the estimates of CMM, the comovement of volatility in SPF forecast updates was high across quarterly forecast horizons extending four periods ahead. In this paper, we impose the common volatility specification in part to reduce dimension in a larger model and in part to improve estimation efficiency for longer forecast horizons for which only annual forecasts are available and even then not at all forecast origins. In addition, when the model is estimated for SPF predictions of a scalar outcome variable, y_t, commonality in uncertainty across forecast horizons appears to be a suitable assumption. By imposing the common-volatility specification on forecast-update gaps, $\tilde{\boldsymbol{\eta}}_t$, instead of forecast updates, $\boldsymbol{\eta}_t$, we shut down the effects of SV on long-horizon forecasts (at horizon H). For forecasts at horizons close to H, our estimates of the corresponding variance coefficients in the bottom right part of $\tilde{\boldsymbol{\Sigma}}$ are also relatively small, which places relatively stronger effects of SV on near- to medium-term forecast horizons.

To simplify the treatment or identification of variable scales, we specify the log SV process to have a mean of 0, reflected in the AR(1) process of equation (8) that omits an intercept, with slope coefficient δ to be estimated. The time-varying variance with mean of 1 scales up a full variance-covariance matrix for the forecast-update gaps, $\tilde{\boldsymbol{\Sigma}}$. Although the ordering of variables commonly affects estimates of VARs with SV processes for each variable (see discussions in studies such as Arias, Rubio-Ramirez, and Shin (2022)), with our common SV specification, the ordering of variables in the model has no impact on estimates. In results omitted in the interest of brevity, we have also considered an extension of the model that allows for additive outlier shocks with the t-distributed formulation of Antolín-Díaz, Drechsel, and Petrella (2021).[19] Although this version of the model attributed some of the large forecast changes during the pandemic to outliers, the overall

[18]Other studies that decompose SPF forecasts into (perceptions of) permanent and transitory components include Krane (2011) and Clements (2022).

[19]The additive outliers serve to lower the correlations among forecast updates for different forecast horizons. In contrast, the multiplicative outliers used by Carriero, et al. (2021) in vector autoregressions with stochastic volatility lower the persistence of volatility changes.

results from the model were essentially the same as those reported for the baseline specification.

4.5 Non-MDS generalization

To allow for possible biases and persistence in forecast errors and expectational updates, we also consider an extension of our model that does not rest on the MDS assumption. In this extended model, we allow for differences to emerge between model-implied forecasts, $E_t y_{t+h}$, and the term structure of SPF-consistent expectations, $y_{t+h|t}$, that is tracked by the state vector $\boldsymbol{Y_t}$.[20] Specifically, in this model, we continue to treat the long-run forecast as a random walk such that μ_t^* is mean 0, but we let $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ follow a VAR(1):[21]

$$\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{\eta}}_{\boldsymbol{t-1}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}. \tag{9}$$

For the VAR shocks, we assume an SV specification that is analogous to what was used for $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ in the non-MDS case, $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \sim N\left(\boldsymbol{0}, \lambda_t \cdot \tilde{\boldsymbol{\Sigma}}\right)$, and we also continue to assume $\mu_t^* \sim N(0, \sigma_*^2)$.

In the non-MDS case, we still rely on the recursive representation of $\boldsymbol{Y_t}$ as driven by a sequence of forecast updates in equation (3). Without the strict MDS assumption, equation (3) can be viewed as a prewhitening step that at least reduces (but does not fully eliminate) persistence in the data. Moreover, when $\tilde{\boldsymbol{G}} = \boldsymbol{0}$, the model nests the MDS case. By centering a Bayesian prior around $\tilde{\boldsymbol{G}} = \boldsymbol{0}$, estimation of the non-MDS model can embody prior beliefs over the (squared-loss) efficiency of the SPF. Indeed, in our Bayesian estimation, we place such a prior (with Minnesota-style shrinkage) on $\tilde{\boldsymbol{G}}$. Critically, forecasts of y_t derived from the non-MDS model generally differ from the SPF, as the non-MDS model allows for serially correlated forecast updates, and thus predictable SPF forecast errors. Differences between model estimates of $E_t y_{t+h}$ and the term structure of SPF-consistent forecasts, measured by $y_{t+h|t}$, reflect the model's assessment of bias in the SPF.

[20] We refer to the set of predictions, $y_{t+h|t}$, contained in $\boldsymbol{Y_t}$ as the term structure of "SPF-consistent" expectations, since $\boldsymbol{Y_t}$ has to be (at least partially) interpolated from observed SPF forecasts for various fixed horizons and events.

[21] Lag order larger than one could also be considered. However, CMM already found little support for such choices.

4.6 Homoskedastic specification

While our baseline model features stochastic volatility, we include in our empirical analysis comparisons to estimates from a model simplified to treat the innovation vector $\boldsymbol{\eta}_t$ as conditionally homoskedastic. In this specification, referred to as CONST below, $\boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Sigma})$.[22]

4.7 Estimation

We estimate the models using Bayesian Markov chain Monte Carlo (MCMC) methods — specifically, a Gibbs sampler. The model estimation is conditioned on joint data for observed realizations and SPF predictions for a given economic variable (like GDP growth), but estimated separately for different economic variables.

In the baseline specification, the objects to be estimated include the δ and σ_ν^2 parameters of the SV process, the constant innovation covariance matrix $\tilde{\boldsymbol{\Sigma}}$, the variance of innovations to the long-run forecast σ_*^2, the time series of the latent volatility state λ_t, and the time series of latent forecast states contained in $\boldsymbol{Y_t}$. We sample the parameters of the SV process using a conventional Gaussian prior and conditional posterior for δ and a standard inverse Gamma prior and conditional posterior for σ_ν^2. We sample $\tilde{\boldsymbol{\Sigma}}$ with an inverse Wishart prior and conditional posterior. We draw σ_*^2 with an inverse Gamma prior and conditional posterior. We estimate the volatility state λ_t with the mixture approach of Kim, Shephard, and Chib (1998), as refined in Omori, et al. (2007).

Sampling the latent forecast states contained in $\boldsymbol{Y_t}$ involves more complexity, and we use a new precision-based sampler to efficiently tackle these complexities. Originally developed in Chib and Jeliazkov (2006) and then Chan and Jeliazkov (2009), precision-based samplers have been used in VAR settings in studies such as Chan (2020). However, in our setting, in which our model assumes

[22] Compared to the SV case, the CONST specification of the non-MDS model also allows for persistent changes in the long-run forecast, since μ_t^* is included in the VAR. We have also considered a non-MDS generalization including conditional homoskedasticity, in which $\boldsymbol{\eta}_t$ follows a VAR(1) with an innovation vector $\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma})$ and transition matrix $\boldsymbol{G}$. With homoskedasticity imposed, results for the non-MDS generalization compared to the MDS (CONST) specification are similar to those reported for specifications featuring SV.

no measurement error, we face an ill-defined posterior precision, $|\operatorname{Var}\left(\boldsymbol{Y_t}|\boldsymbol{Z}^t\right)| = 0$, which cannot be directly handled by conventional precision-based samplers. To efficiently draw $\boldsymbol{Y}^t|\boldsymbol{Z}^t$ in this case, we use a new precision-based sampler developed and detailed in Mertens (2022).[23] We retain 3,000 draws after a burn-in sample of 3,000 initial draws. When simulating the model's predictive density we sample 100 paths of future realizations of stochastic volatility and other state variables for each MCMC draw, resulting in $S = 300,000$ predictive density draws.

5 Results

This section first documents the historical time variation in the volatility of SPF forecast errors. We then present estimates of the term structures of expectations using our baseline model, followed by estimates of the term structures of uncertainty. It next compares forecast accuracy for our baseline model with SV that exploits MDS assumptions to a model without SV and a model that does not include MDS assumptions. Finally, the section presents a practical application using our estimated term structure of forecasts: constructing fan charts with calendar-year (four-quarter growth rates or fourth-quarter levels) forecasts, in the style used by the FOMC.

5.1 Time variation in volatility

To show a key factor in our rationale for including stochastic volatility in the model, Figure 1 provides time-varying quarterly volatility estimates obtained with the expectational updates for GDP growth for selected horizons, including the short horizons of $h = 0, 3$ for which the expectational updates are directly observable and longer horizons of $h = 7, 11$ (corresponding to two and three years ahead) for which quarterly expectational updates are not directly observable and are instead inferred from the model and observed annual forecasts.[24] Specifically, the black lines

[23] Alternatively, the sampling of the latent state vector can be handled with a simulation smoother such as that of Durbin and Koopman (2002), albeit at slightly higher computational costs.

[24] The directly observable elements of $\boldsymbol{\eta}_t$ for $h = 0, \ldots, 4$ were also used as data in the CMM model.

provide the full-sample (smoothed) estimates of time-varying volatility.[25] While not shown in the interest of chart readability, real-time estimates of stochastic volatility (obtained by looping over time and estimating a historical volatility path at each forecast origin) that underlie the forecast results considered in the next section are very similar to the full-sample estimates. For comparison to the volatility estimates, the figures include (in grey bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the volatility estimates.

[Figure 1 about here.]

Consistent with the findings of CMM, the volatility estimates display several broad features (although we only show estimates for GDP growth, estimates for other variables share the same features; the supplementary online appendix reports full-sample estimates of the scale volatility factor for all variables). First, the time variation in volatility is considerable. In the period preceding the temporary surge of volatility induced by the COVID-19 outbreak, the highs in the volatility estimates are typically 3 to 4 times the levels of the lows in the estimates. As documented in other studies, such as Carriero, et al. (2021), volatility soared to unprecedented levels in the first few quarters of the pandemic. Second, some of the time variation occurs at low frequencies, chiefly with the Great Moderation of the 1980s. But some of the time variation is cyclical, as volatility has some tendency to rise temporarily around recessions. For example, the volatility of GDP growth rises with most recessions. Third, by construction, the volatility estimates obtained from our common-SV specification display perfect comovement across horizons (except for H). However, the observable data for elements of $\tilde{\boldsymbol{\eta}}_t$ share this commonality.

5.2 Term structures of expectations

To illustrate the term structures of expectations produced by our model, Figures 2-5 present estimates of point forecasts for, respectively, GDP growth, the unemployment rate, CPI inflation, and the T-bill rate, using SPF forecasts available as of the indicated forecast origins. As detailed

[25] The time-varying volatilities reflect the diagonal elements of $\mathrm{Var}_t\left(\boldsymbol{\eta}_t\right)$, which are computed from equation (7).

above, the model uses as inputs quarterly SPF forecasts up through the 4-steps-ahead horizon and the available annual fixed-event forecasts (except for the current year). With fixed-event forecasts available up to three years ahead, the model yields estimated quarterly forecasts through the 15-steps-ahead horizon for GDP growth, the unemployment rate, and the T-bill rate. For CPI inflation, the longest-available horizon for calendar-year forecasts is two years, from which we extract term structures up to $H = 11$ quarters out.[26] In each panel, red diamonds give the available quarterly fixed-horizon forecasts from the SPF, and dotted lines give the available annual fixed-event forecasts that bear on the indicated quarter, with next year in blue, two years ahead in green, and three years ahead in red. Because the annual forecasts of GDP growth are percent changes in annual averages of GDP, the annual forecasts reflect information in quarterly growth rates for 7 quarters, as indicated in our previous discussion of the model's measurement equations, in equation (6); in other cases, the annual forecasts reflect information in forecasts for 4 quarters. The black lines provide the real-time quarterly SPF-consistent forecast estimates, with 68 percent credible sets for their latent values (not the realized outcomes) indicated with the thin lines.

[Figure 2 about here.]

One pattern in the results reflects the design of the model: The model yields quarterly forecasts for horizons 0 through 4 quarters that match exactly the reported SPF projections, with no uncertainty around the estimate. Only the model-based forecasts for subsequent quarters show any uncertainty around them. The second notable pattern in the results is that the estimated quarterly forecasts for horizons of 5 through 15 quarters ahead interpolate through the annual fixed-event forecasts, with some fluctuations around the annual forecasts. For example, in Figure 2's results

[26]The H-steps-ahead forecasts, $y_{t+H|t}$ in (2), constitute what Kozicki and Tinsley (2001) called the shifting endpoints of the term structure of expectations. The supplementary online appendix reports estimates of these shifting endpoints over time. For GDP growth, CPI inflation, and the T-bill rate, the SPF collects 10-year-ahead average expectations, which are tracked closely by our endpoint estimates for term structures. In addition, the supplementary online appendix considers estimates generated by an extended model that conditions on the SPF's 10-year-ahead expectations for GDP growth, CPI inflation, and T-bill returns, respectively.

for GDP growth, at horizons of 5 through 15 quarters, the black lines representing quarterly forecasts generally intersect the dots (representing the annual forecasts) within the first 3 or 4 quarters covered. In 2009Q1, when the SPF reported a next-year annual forecast but not subsequent year forecasts, the model-based quarterly forecast intersects the annual forecast before the mid-point of the blue dots, and then remains above the 2009 annual year forecast for subsequent quarters. Similarly, in 2019Q4 (note that this panel shows dots for two rather than three years of annual forecasts, because due to strong overlap with quarterly forecasts, the annual forecast from the SPF for the calendar year of 2020 is not used in model estimation), the model-based quarterly forecast intersects the annual forecast for 2022 before the mid-point of the red dots, and then remains above the 2022 annual forecast for a few quarters before moving to a lower level. In a number of the other cases shown, the model's interpolated quarterly forecasts of GDP growth intersect the annual forecast a second time late in the period covered by the annual prediction. This pattern applies to two of the three years covered in the 2009Q2 results, all three years in the 2017Q3 estimates, and one of the two years in the 2019Q4 results. Overall, while the model's estimates of GDP growth forecasts for quarters 5 through 15 interpolate through the available annual forecasts, they display some fluctuations.

Note that some of the fluctuations in forecasts that occur from quarters 5 through 15 can be attributed to the dynamics in the SPF's quarterly point forecasts at shorter horizons and simple accounting between those forecasts and the longer-horizon annual forecasts. That is, some fluctuations stem directly from the reported SPF forecasts and are not necessarily artificial wiggles driven by the model. The results for 2009Q2 in the upper right panel of Figure 2 provide a good example. In this case, for the next year (2010), quarterly forecasts are only unobserved for 2010Q3 and 2010Q4. With the published forecasts for earlier quarters (red dots) below the annual forecast represented by the blue dots, the forecasts for the second half of 2010 have to lie above the annual rate. Arithmetically, because these forecasts for 2010Q3 and 2010Q4 also need (to match the 2010 annual forecast) to be above the annual forecast for 2011 represented by the green dots, some later quarterly forecasts have to be lower than the annual forecast for 2011. For similar accounting reasons, the term structures of expectations in some other examples — such as GDP growth forecasts in 2019Q4 shown in the lower right panel — do not have the kind of fluctuations seen in the case of GDP growth forecasts in 2009Q2. Related, in results detailed in the supplementary online appendix, it is possible to derive univariate processes for the time series y_t implied by estimates of the state-space model (conditioning on the entire SPF term structure). Essentially, using the model's steady-state Kalman filter and the Kalman gain, we can obtain moving average (MA) coefficients of the process for y_t. In estimates for GDP growth and the two inflation measures, the coefficient profiles resemble low-order MA processes, whereas for the unemployment and T-bill rates the patterns resemble more persistent AR processes. Fluctuations in quarterly forecasts across horizons can be seen as consistent with these implied univariate processes.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

Estimates for the unemployment rate, CPI inflation, and the T-bill rate in Figures 3–5 are qualitatively similar. Results for the unemployment and T-bill rates show two other patterns that either differ modestly from or are not evident in the GDP and CPI results. First, the annual forecasts are simply annual averages of forecasts within the 4 quarters of the year, and the model's interpolation of quarterly forecasts tends to show less variation. For horizons of 5 quarters and beyond, the estimated quarterly forecasts typically intersect the annual forecasts once, not twice. Second, with the availability of annual forecasts at longer horizons more mixed for the unemployment and T-bill rates, the estimated term structures of forecasts for these variables indicate that the availability of annual forecasts at longer horizons tends to increase the precision of the quarterly forecast estimates. That is, the quarterly estimates tend to be more precise when annual forecasts for the quarters are available than when they are not available. As examples, in the upper left panel's results for a forecast origin of 2009Q1, the available fixed-event annual forecasts were limited to the next year; at that time, the SPF did not publish forecasts for subsequent years. For 2009Q2 or 2009Q3, covered in the upper right panel, the SPF included two additional years of annual forecasts of the unemployment and T-bill rates. Compared to estimates for other origins, the results for 2009Q1 show the credible sets around the forecast estimates at horizons of 6 quarters or more to be relatively wide. The results for CPI expectations, shown in Figure 4, are an exception to the latter pattern. For CPI, the SPF has provided calendar-year forecasts for up to three years out since 2005, which allows us to inform estimates of the term structure of CPI expectations throughout the run-up of the Great Recession as shown in the figure.

5.3 Term structures of uncertainty

After illustrating the basic properties of our model regarding the point forecasts, we now add forecast uncertainty to the picture. Fan charts of quarterly forecasts provide one reading on forecast uncertainty and its time variation; for brevity, and partly because below we will examine fan charts for calendar-year rather than quarterly forecasts, the supplementary online appendix reports fan charts of SPF-consistent quarterly forecasts for horizons through 15 quarters ahead, including the point forecast and the predictive density as captured by 68 percent bands. These results show that the width of the uncertainty bands appears to vary over time (across forecast origins) and, as expected, forecast uncertainty tends to rise with the forecast horizon.

[Figure 6 about here.]

To directly assess changes in uncertainty across horizons and over time, Figure 6 depicts the term structure of uncertainty around quarterly forecasts, from 1983 to 2021. For constructing the figure, uncertainty is measured by the width of the 68 percent bands of the model's predictive densities estimated in real time. For readability, the chart includes a subset of quarterly horizons, including a few short ones and longer horizons at increments of 3 or 4 quarters (for the inflation measures, because the available annual forecasts from the SPF omit longer horizons, the horizons included in the figure are shorter than those for growth and the unemployment and T-bill rates).

In these results, uncertainty is noticeably higher at longer horizons (8 or more quarters) than shorter horizons (0 to 4 quarters). From 0 to 4 quarters, uncertainty gradually increases. From 8 to 15 quarters, uncertainty continues to rise, but typically by less than in the short horizon case. Consistent with the patterns in full-sample SV estimates for GDP growth discussed above, the uncertainty of out-of-sample forecasts fluctuates significantly over time. For GDP growth, uncertainty moderated in the early years of the sample but then rose some following the 2001 recession and more notably around the Great Recession and again a few years into the ensuing recovery. Then the outbreak of COVID-19 produced an unprecedented, but temporary, spike in uncertainty in 2020. The uncertainty estimates for the other variables also decline significantly with the Great Moderation and then fluctuate over the remainder of the sample.

5.4 Forecast performance: Contributions of stochastic volatility

Building on the results above that show the stochastic volatility component of the model to capture or produce significant time variation in forecast error variances and forecast uncertainty, we conduct an out-of-sample evaluation of real-time forecasts generated by the baseline and CONST versions of the model. Table 2 compares the accuracy of point forecasts, evaluated in terms of RMSE, as well as density predictions, evaluated by the continuous ranked probability score (CRPS). The results are reported as ratios with SV results in the denominator, so that entries above (below) 1 indicate the SV forecasts to be more (less) accurate than CONST forecasts. From 1983Q3 onward, out-of-sample forecasts are generated for all quarterly horizons from $h = 0$ to 15, based on all available data since 1968Q4, by re-estimating the model at each forecast origin and simulating its predictive density. In these results, the sample uses data through 2022Q2; results in the supplementary online appendix for a sample through 2016Q4 that omits pandemic effects are qualitatively similar, with one modest difference noted below.

[Table 2 about here.]

Over the full sample, the RMSE ratios are generally close to 1, indicating that the point forecasts from the SV and CONST specifications are very similar. This result may be expected, given that the point forecasts from both models interpolate between the available annual point forecasts from the SPF. In a handful of cases (mainly PGDP and TBILL forecasts at longer horizons), the SV-based forecasts are slightly more accurate than the CONST forecasts. In some others, the SV-based forecasts are slightly less accurate.

As may also be expected, modestly sharper differences in forecast accuracy across the specifications are evident in the CRPS ratios, which capture density accuracy.[27] For GDP growth and PGDP inflation, the SV forecasts are significantly more accurate than the CONST forecasts, with gains that peak at 7 percent for GDP growth and 12 percent for PGDP inflation. SV offers less advantage over the CONST specification in the results for unemployment and the T-bill rate. However, in results in the supplementary online appendix, when the pandemic is omitted from the evaluation sample, SV offers more consistent small advantages over the CONST specification for these variables.

5.5 Forecast performance without MDS assumption

So far, we have considered model estimates that treat the SPF point forecasts as optimal forecasts, and embody a martingale difference assumption regarding forecast updates. As described in Section 4.5, our model can also be extended to treat SPF predictions as martingales in only the model's Bayesian prior while allowing for VAR dynamics in the posterior estimates of the forecast update vector $\boldsymbol{\eta}_t$. Accordingly, the estimated model allows the data to speak on the extent to which historical bias in observed SPF forecasts, resulting in persistent forecast errors, is sufficient to push the estimates away from the MDS baseline. Comparing the accuracy of forecasts across the models sheds light on the potential benefits to forecast accuracy that could be achieved by allowing

[27]Although a number of studies of time series models have found modestly larger density gains associated with SV, these studies typically commingle benefits to point forecasts with benefits to the variance aspect of the density forecasts. In our case, the point forecasts are anchored around the available quarterly and annual SPF predictions, so any gains in density accuracy come more from variance-related aspects of the forecast distribution than differences in point forecasts.

time-varying bias in SPF forecasts.

[Table 3 about here.]

Accordingly, Table 3 reports the results of an out-of-sample evaluation of real-time forecasts generated by the MDS and non-MDS versions of the model. In the interest of brevity and given our preference for SV over the CONST model, we compare MDS and non-MDS specifications for just the SV case. As earlier, we consider point forecasts, evaluated in terms of RMSE, as well as density predictions, evaluated by the CRPS, for a sample from 1983Q3 through 2022Q2.

As indicated in the left panel of the table, the accuracy of point forecasts from the non-MDS specifications is generally quite similar to the accuracy of their MDS-based counterparts. In particular, for GDP growth, the RMSE ratios are 1.00 at all horizons except for $h = 1$. Similarly, for the unemployment rate, the RMSE ratios are 1.00 for most horizons and 1.01 for the others. For the inflation measures, the non-MDS specification slightly improves on the MDS baseline at some horizons. The relative performance of the specifications is more varied across horizons for the T-bill rate, with the non-MDS forecasts slightly more accurate than the baseline at shorter horizons and less accurate at long horizons. Over the full sample, the CRPS results in the right panel of the table indicate that the density accuracy of the two specifications shows the same patterns as the point forecast accuracy.

[Figure 7 about here.]

The statistics of average forecast performance in Table 3 suggest little effect of modeling bias in SPF-consistent expectations. Instead, the non-MDS model picks up on episodic increases in bias (in absolute value), in particular during recessions. Nevertheless, these episodic increases are not very persistent, and tend to be overshadowed by much larger increases in forecast error variance. For selected variables and forecast horizons, Figure 7 reports real-time estimates of bias, $b_{t+h|t} = E_t y_{t+h} - y_{t+h|t}$, and ex-ante MSE of SPF-consistent forecasts, $\text{MSE}_{t+h|t} = E_t \left(y_{t+h} - y_{t+h|t} \right)^2$. E_t denotes the expectations operator of the non-MDS model, and $y_{t+h|t}$ denotes the (posterior mean) estimate of the SPF-consistent expectations at horizon h as tracked by the model's state vector $\boldsymbol{Y_t}$. All expectations are obtained as end-of-sample (real-time) estimates generated from the MCMC runs over growing samples also used in our analysis of average forecast performance reported above. The left-hand column of Figure 7 shows the evolution of bias estimates, $b_{t+h|t}$, over time, with short-lived but very visible increases in amplitude in all recessions over our evaluation window starting in 1986.[28] The right-hand column of the figure reports a decomposition of MSE into the sum of squared bias and forecast error variance, $\text{MSE}_{t+h|t} = \text{Var}_t\left(y_{t+h}\right) + b^2_{t+h|t}$ (where Var_t denotes the model's conditional variance operator), which shows that any bias in SPF-consistent expectations is generally dwarfed — i.e., bias can barely be distinguished in the right-hand column — by unavoidable forecast error variance, in particular during recessions.

Overall, these results suggest little benefit to generalizing our baseline model to allow departures from our MDS assumption. But they also show that there is little cost in forecast accuracy to doing so. Accordingly, our results are consistent with the prior literature that, on one hand, reports considerable serial correlation in forecast updates of profession forecasters, like the SPF; see, for example, Coibion and Gorodnichenko (2015) and Coibion, Gorodnichenko, and Kamdar (2018). But, on the other hand, the literature has also found such deviations from SPF forecast efficiency to be episodic and hard to consistently exploit in real time, as noted, for example, by Croushore (2010), and more recently by Foerster and Matthes (2022) and Hajdini and Kurmann (2022). Our results on any biases being short-lived, contributing at most only a small fraction to overall MSE, are consistent with notions of SPF forecasts providing at least boundedly optimal, if not highly rational, forecasts, as discussed in, for example, Coibion, Gorodnichenko, and Kamdar (2018) and Mertens and Nason (2020). For our purposes, these results motivate our interest in centering estimates of the term structure of expectations and uncertainty around the SPF. We acknowledge, however, that Bianchi, Ludvigson, and Ma (2022) have recently provided evidence that machine learning methods show more capability to improve on the accuracy of survey-based forecasts.

[28]Results for the COVID-19 recession of 2020 are not shown for reasons of scale, but are provided in the supplementary online appendix.

5.6 Practical applications using term structure of forecasts

As a practical matter, our model of the term structure of SPF forecasts can be used to produce SPF-consistent fan charts directly analogous to those published by central banks. In some economies, central bank fan charts refer to four-quarter (year-over-year) growth rates, reported quarter by quarter. For example, since 2017 the US Federal Reserve publishes fan charts as part of the FOMC's Summary of Economic Projections (SEP) that report calendar-year forecasts that are measured as Q4/Q4 growth rates of GDP and prices and the Q4 level of the unemployment rate. The width of the uncertainty bands in these SEP fan charts is two times the historical RMSEs that are reported in the SEP's Table 2, obtained as average RMSEs for a few different forecasts computed over 20-year rolling windows, as developed in Reifschneider and Tulip (2019). From 2007 through 2016, the SEP did not report fan charts, but it included historical RMSEs in its Table 2, with the intention of informing the FOMC participants' assessments of uncertainty.[29]

[Figure 8 about here.]

To illustrate, Figures 8 and 9 provide SEP-style fan charts with calendar-year forecasts of GDP growth, CPI inflation, and the unemployment rate, for forecast origins of 2011Q1, 2012Q1, and 2014Q1. As detailed below, this period was associated with some changes in the FOMC's subjective assessments of uncertainty. Using estimates from our baseline SV model, the charts provide SPF-consistent point forecasts and 68 percent uncertainty bands.[30] We produced the SPF-consistent calendar-year forecasts by taking the appropriate transformations of the quarterly forecasts through H=15 from our estimated term structure of expectations.[31]

Across these origins, the SPF's point forecasts of growth were fairly flat across horizons of this application, with uncertainty gradually rising with the horizon. As we will detail below, for GDP

[29] Historical RMSEs were provided as part of the SEP's Table 2 since inception of the SEP format in October 2007; see `https://www.federalreserve.gov/monetarypolicy/files/fomcminutes20071031.pdf`.

[30] Q1 SPF forecasts are published in mid-February of each year. What we treat as the Q1 SEPs were published in late January (2011 and 2012) or mid-March (2014).

[31] To be more precise, for each posterior draw of a sequence of quarterly forecasts, we transformed to the calendar-year forecasts of interest and then tabulated the reported means and credible sets.

growth and the unemployment rate, our estimates of uncertainty (the width of the bands in these charts) were well below the estimates used in the SEP's fan charts. The SPF's point forecasts of the unemployment rate declined significantly across horizons in 2011Q1 and 2012Q1 and edged down more modestly in 2014Q1, in all cases with uncertainty gradually rising with the horizon. For each of these forecast origins, the SPF's point forecasts of CPI inflation showed small gradual increases across horizons, accompanied by uncertainty rising with the horizon.

[Figure 9 about here.]

Figure 9 directly compares over time uncertainty as estimated from our model (with the width of 68 percent bands) to uncertainty measured as the SEP currently does, as two times the historical RMSEs published in the SEP's Table 2. For each year from 2008 through 2021, we report estimates for SPF and SEP forecasts published in the first quarter.[32] As might be expected, our estimates show more variation over time than do the SEP measures, because the model includes SV, whereas the SEP measures treat forecast error variances as being constant over 20-year windows. Following the volatility induced by the Great Recession, our estimates of uncertainty rise relatively quickly and sharply; the SEP measures eventually reflect the effects of the large forecast errors of the recession — more clearly so for growth and unemployment than CPI inflation — but rise relatively slowly and by less than the SV estimates. But our SV-based estimates fell quickly. In the 2011Q1, 2012Q1, and 2014Q1 examples, SPF-consistent forecasts from our model imply noticeably less uncertainty in the outlook for GDP growth than the SEP-based approach. The same pattern applies to the unemployment rate examples, except that, in 2010Q1, our estimates of uncertainty were relatively similar to the SEP-based measures. From 2015 to 2019, both our estimates and the SEP estimates are relatively stable, with the latter continuing to exceed the former. Following the volatility induced by the outbreak of the pandemic in 2020, our SV-based estimates of uncertainty rose sharply in 2021Q1, to exceed the SEP estimates. In the case of CPI inflation, since 2016 the

[32]The available time series for the SEP begins in October 2007, making the first available Q1 observation 2008. The charts omit SEP observations for 2020 because, in March 2020, in the aftermath of the pandemic's outbreak, the FOMC did not publish an SEP.

SPF estimates of forecast uncertainty have tended to be a little lower than the SEP estimates but are broadly similar in their level.

In addition to providing fan charts and historical RMSEs to quantify uncertainty around the outlook, the SEP includes measures of FOMC participants' subjective assessments of uncertainty. In the current SEP, a note in the fan charts explains: "Because current conditions may differ from those that prevailed, on average, over the previous 20 years, the width and shape of the confidence interval estimated on the basis of the historical forecast errors may not reflect FOMC participants' current assessments of the uncertainty and risks around their projections...." Since December 2020, the SEP has included in Figure 4.D a chart of a diffusion index that summarizes the Committee's subjective assessments of uncertainty, using information from additional survey questions.

The period from 2011 to 2014 provides examples of the SEP's subjective assessments of uncertainty differing from historical quantitative assessments. In our model estimates, uncertainty was relatively stable from 2011 and 2012 to 2014 (although inflation uncertainty drifted down a bit). While the levels of estimates for GDP growth and unemployment based on the SEP's current treatment of fan charts were generally higher than our estimates, the SEP-type measures also showed little change across these three years, except that GDP growth uncertainty rose from 2012 to 2014. In contrast, the SEP's subjective assessments show some relatively sizable changes over the period. From 2012 to 2014, the subjective uncertainty assessments around growth, unemployment, and inflation shifted from "elevated" to "normal," reflected in a drop in the diffusion index.

6 Conclusion

For application to survey forecasts such as the SPF, this paper develops a model that permits the estimation of a term structure of both expectations and forecast uncertainty. In broad terms, our model can be seen as a multivariate unobserved components model with stochastic volatility. Our approach exactly replicates a given set of predictions from the SPF or a similar forecast source without measurement error. Extending previous work, our model includes not only fixed-horizon forecasts but also fixed-event forecasts, including at horizons beyond the fixed-horizon maximum, and it accommodates variation over time in the available horizons of fixed-event forecasts.

Our empirical results first establish that our model yields quarterly forecasts that perfectly match and interpolate through the annual fixed-event point forecasts. We then show that the model's estimates of uncertainty vary over time and generally rise with the forecast horizon. Finally, we apply our approach to constructing fan charts with SPF-based calendar-year (four-quarter growth rates) forecasts like those published by the FOMC. In recent years, our model applied to SPF forecasts yields estimates of uncertainty around GDP growth and unemployment forecasts that are lower than those implied by the historical forecast RMSEs that underlie the FOMC's fan charts. Our approach could be used in real time with each new release of the SPF to produce updated fan charts of point forecasts and estimates of forecast uncertainty.

We leave some possible extensions of our framework as topics for future research. For example, forecasts from individual SPF participants could be integrated into the analysis. Using individual forecasts in our setting would require some extensions to accommodate their available shorter samples with more missing observations. Such an extension might also rest on analytics that more formally develop a relationship between aggregate forecasts and disagreement at the individual level, for example, by a model of dispersed information as in Lahiri and Sheng (2010) or Binder, McElroy, and Sheng (2022). Alternatively, following Cakmakli and Demircan (2022), disagreement measures might be used as observable signals of uncertainty in measurement equations for otherwise latent SV process(es) of the model. In a similar vein, measures of financial stress might help sharpen the estimates of uncertainty, building on the results of Adams, et al. (2021), who found SPF forecast errors susceptible to the kind of downside vulnerabilities reported by Adrian, Boyarchenko, and Giannone (2019). Finally, our framework could be extended to jointly model survey-implied term structures for multiple variables (like inflation and output) and identify structural shocks, for example, as in Benhima and Poilly (2021) or Doh and Smith (2022).[33]

[33]CMM considered joint estimation of their model for several economic variables, but without consistent gains in forecast performance or other notable changes in estimates.

References

Adams, Patrick A., Tobias Adrian, Nina Boyarchenko, and Domenico Giannone (2021), "Forecasting macroeconomic risks," *International Journal of Forecasting*, 37, 1173–1191, https://doi.org/10.1016/j.ijforecast.2021.01.003.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019), "Vulnerable growth," *American Economic Review*, 109, 1263–1289, https://doi.org/10.1257/aer.20161923.

Ang, Andrew, Geert Bekaert, and Min Wei (2007), "Do macro variables, asset markets, or surveys forecast inflation better?" *Journal of Monetary Economics*, 54, 1163–1212, https://doi.org/10.1016/j.jmoneco.2006.04.006.

Antolín-Díaz, Juan, Thomas Drechsel, and Ivan Petrella (2021), "Advances in nowcasting economic activity: Secular trends, large shocks and new data," Discussion Paper 15926, CEPR.

Arias, Jonas E., Juan F. Rubio-Ramirez, and Minchul Shin (2022), "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," *Journal of Econometrics*, forthcoming, https://doi.org/10.1016/j.jeconom.2022.04.013.

Aruoba, S. Boragan (2020), "Term structures of inflation expectations and real interest rates," *Journal of Business & Economic Statistics*, 38, 542–553, https://doi.org/10.1080/07350015.2018.1529599.

Banbura, Marta, Federica Brenna, Joan Paredes, and Francesco Ravazzolo (2021), "Combining Bayesian VARs with survey density forecasts: Does it pay off?" Working Paper Series 2543, European Central Bank.

Bassetti, Federico, Roberto Casarin, and Marco Del Negro (2022), "Inference on probabilistic surveys in macroeconomics with an application to the evolution of uncertainty in the Survey of Professional Forecasters during the COVID pandemic," in *Handbook of Economic Expectations* eds. by Wilbert van der Klaauw, Giorgio Topa, and Ruediger Bachmann: Elsevier.

Benhima, Kenza, and Céline Poilly (2021), "Does demand noise matter? Identification and implications," *Journal of Monetary Economics*, 117, 278–295, https://doi.org/10.1016/j.jmoneco.2020.01.006.

Beveridge, Stephen, and Charles R. Nelson (1981), "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "Business Cycle"," *Journal of Monetary Economics*, 7, 151–174, https://doi.org/10.1016/0304-3932(81)90040-4.

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma (2022), "Belief distortions and macroeconomic fluctuations," *American Economic Review*, 112, 2269–2315, https://doi.org/10.1257/aer.20201713.

Binder, Carola, Tucker S. McElroy, and Xuguang S. Sheng (2022), "The Term Structure of Uncertainty: New Evidence from Survey Expectations," *Journal of Money, Credit and Banking*, 54, 39–71, https://doi.org/10.1111/jmcb.12811.

Cakmakli, Cem, and Hamza Demircan (2022), "Using survey information for improving the density nowcasting of US GDP," *Journal of Business & Economic Statistics*, forthcoming, https://doi.org/10.1080/07350015.2022.2058000.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016), "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390, https://doi.org/10.1080/07350015.2015.1040116.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2021), "Addressing COVID-19 outliers in BVARs with stochastic volatility," *Review of Economics and Statistics*, forthcoming, https://doi.org/10.1162/rest_a_01213.

Chan, Joshua C. C. (2020), "Large Bayesian VARs: A flexible Kronecker error covariance structure," *Journal of Business & Economic Statistics*, 38, 68–79, https://doi.org/10.1080/07350015.2018.1451336.

Chan, Joshua C. C., and Ivan Jeliazkov (2009), “Efficient simulation and integrated likelihood estimation in state space models,” *International Journal of Mathematical Modelling and Numerical Optimization*, 1, 101–120, https://doi.org/10.1504/IJMMNO.2009.03009.

Chib, Siddhartha, and Ivan Jeliazkov (2006), “Inference in semiparametric dynamic models for binary longitudinal data,” *Journal of the American Statistical Association*, 101, 685–700, https://doi.org/10.1198/016214505000000871.

Clark, Todd E. (2011), “Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility,” *Journal of Business and Economic Statistics*, 29, 327–341, https://doi.org/10.1198/jbes.2010.09248.

Clark, Todd E., Gergely Ganics, and Elmar Mertens (2022), “What is the predictive value of SPF point and density forecasts?”, *mimeo*.

Clark, Todd E., Michael W. McCracken, and Elmar Mertens (2020), “Modeling time-varying uncertainty of multiple-horizon forecast errors,” *The Review of Economics and Statistics*, 102, 17–33, https://doi.org/10.1162/rest_a_00809.

Clark, Todd E., and Francesco Ravazzolo (2015), “Macroeconomic forecasting performance under alternative specifications of time-varying volatility,” *Journal of Applied Econometrics*, 30, 551–575, https://doi.org/10.1002/jae.2379.

Clements, Michael P. (2018), “Are macroeconomic density forecasts informative?” *International Journal of Forecasting*, 34, 181–198, https://doi.org/10.1016/j.ijforecast.2017.10.004.

——— (2022), “Individual forecaster perceptions of the persistence of shocks to GDP,” *Journal of Applied Econometrics*, 37, 640–656, https://doi.org/10.1002/jae.2884.

Clements, Michael P., and Ana Beatriz Galvão (2017), “Model and survey estimates of the term structure of US macroeconomic uncertainty," *International Journal of Forecasting*, 33, 591–604, https://doi.org/10.1016/j.ijforecast.2017.01.004.

Clements, Michael P., Robert W. Rich, and Joseph Tracy (2022), "Surveys of professionals," in *Handbook of Economic Expectations* eds. by Wilbert van der Klaauw, Giorgio Topa, and Ruediger Bachmann: Elsevier.

Coibion, Olivier, and Yuriy Gorodnichenko (2015), "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 105, 2644–78, https://doi.org/10.1257/aer.20110306.

Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar (2018), "The formation of expectations, inflation, and the Phillips curve," *Journal of Economic Literature*, 56, 1447–1491, https://doi.org/10.1257/jel.20171300.

Croushore, Dean (2010), "An evaluation of inflation forecasts from surveys using real-time data," *The B.E. Journal of Macroeconomics*, 10, 1–32, https://doi.org/10.2202/1935-1690.1677.

Crump, Richard K., Stefano Eusepi, Emanuel Moench, and Bruce Preston (2021), "The term structure of expectations," Staff Reports 992, Federal Reserve Bank of New York.

D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101, https://doi.org/10.1002/jae.1257.

Doh, Taeyoung, and A. Lee Smith (2022), "A new approach to integrating expectations into VAR models," *Journal of Monetary Economics*, 132, 24–43, https://doi.org/10.1016/j.jmoneco.2022.08.001.

Dovern, Jonas, Ulrich Fritsche, and Jiri Slacalek (2012), "Disagreement among forecasters in G7 countries," *Review of Economics and Statistics*, 94, 1081–1096, https://doi.org/10.1162/REST_a_00207.

Durbin, J., and S.J. Koopman (2002), "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603–615, https://doi.org/10.1093/biomet/89.3.603.

Farmer, Leland, Emi Nakamura, and Jon Steinsson (2022), "Learning about the long run," NBER Working Paper 29495, National Bureau of Economic Research, https://doi.org/10.3386/w29495.

Faust, Jon, and Jonathan H. Wright (2009), "Comparing Greenbook and reduced form forecasts using a large realtime dataset," *Journal of Business & Economic Statistics*, 27, 468–479, https://doi.org/10.1198/jbes.2009.07214.

Foerster, Andrew, and Christian Matthes (2022), "Learning about regime change," *International Economic Review*, forthcoming, https://doi.org/10.1111/iere.12585.

Frey, Christoph, and Frieder Mokinski (2016), "Forecasting with Bayesian vector autoregressions estimated using professional forecasts," *Journal of Applied Econometrics*, 31, 1083–1099, https://doi.org/https://doi.org/10.1002/jae.2483.

Ganics, Gergely, Barbara Rossi, and Tatevik Sekhposyan (2021), "From fixed-event to fixed-horizon density forecasts: Obtaining measures of multi-horizon uncertainty from survey density forecasts," *Journal of Money, Credit, and Banking*, forthcoming.

Glas, Alexander, and Matthias Hartmann (2022), "Uncertainty measures from partially rounded probabilistic forecast surveys," *Quantitative Economics*, 13, 972–1022, https://doi.org/10.3982/QE1703.

Grishchenko, Olesya, Sarah Mouabbi, and Jean-Paul Renne (2019), "Measuring inflation anchoring and uncertainty: A U.S. and euro area comparison," *Journal of Money, Credit, and Banking*, 51, 1053–1096, https://doi.org/10.1111/jmcb.12622.

Hajdini, Ina, and André Kurmann (2022), "Predictable forecast errors in full-information rational expectations models with regime shifts," *mimeo*.

Hepenstrick, Christian, and Jason Blunier (2022), "What were they thinking? Estimating the quarterly forecasts underlying annual growth projections," Working Paper 2022-05, Swiss National Bank.

Jo, Soojin, and Rodrigo Sekkel (2019), "Macroeconomic uncertainty through the lens of professional forecasters," *Journal of Business & Economic Statistics*, 37, 436–446, https://doi.org/10.1080/07350015.2017.1356729.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *The Review of Economic Studies*, 65, 361–393, https://doi.org/10.1111/1467-937X.00050.

Kozicki, Sharon, and P. A. Tinsley (2001), "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics*, 47, 613–652, https://doi.org/10.1016/S0304-3932(01)00054-X.

Kozicki, Sharon, and P.A. Tinsley (2012), "Effective use of survey information in estimating the evolution of expected inflation," *Journal of Money, Credit, and Banking*, 44, 145–169, https://doi.org/10.1111/j.1538-4616.2011.00471.x.

Krane, Spencer D. (2011), "Professional forecasters' view of permanent and transitory shocks to GDP," *American Economic Journal: Macroeconomics*, 3, 184–211, https://doi.org/10.1257/mac.3.1.184.

Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo (2017), "Using entropic tilting to combine BVAR forecasts with external nowcasts," *Journal of Business & Economic Statistics*, 35, 470–485, https://doi.org/10.1080/07350015.2015.1087856.

Lahiri, Kajal, and Xuguang Sheng (2010), "Measuring forecast uncertainty by disagreement: The missing link," *Journal of Applied Econometrics*, 25, 514–538, https://doi.org/10.1002/jae.1167.

Mariano, Roberto S., and Yasutomo Murasawa (2003), "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443, https://doi.org/10.1002/jae.695.

Mertens, Elmar (2022), "A precision-based sampler for unobserved component models without measurement error," *mimeo*.

Mertens, Elmar, and James M. Nason (2020), "Inflation and professional forecast dynamics: An evaluation of stickiness, persistence, and volatility," *Quantitative Economics*, 11, 1485–1520, https://doi.org/10.3982/QE980.

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima (2007), "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425–449, https://doi.org/10.1016/j.jeconom.2006.07.008.

Patton, Andrew J., and Allan Timmermann (2011), "Predictability of output growth and inflation: A multi-horizon survey approach," *Journal of Business & Economic Statistics*, 29, 397–410, https://doi.org/10.1198/jbes.2010.08347.

——— (2012), "Forecast rationality tests based on multi-horizon bounds," *Journal of Business & Economic Statistics*, 30, 1–17, https://doi.org/10.1080/07350015.2012.634337.

Reifschneider, David, and Peter Tulip (2019), "Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve's approach," *International Journal* *of Forecasting*, 35, 1564–1582, https://doi.org/10.1016/j.ijforecast.2018.07.016.

Wright, Jonathan H. (2013), "Evaluating real-time VAR forecasts with an informative democratic prior," *Journal of Applied Econometrics*, 28, 762–776, https://doi.org/10.1002/jae.2268.

Table 1: Availability of SPF point forecasts

		Fixed-horizon	Fixed-event calendar years		
Variable	Mnemonic	Quarters 0 – 4	next	2-year	3-year
Real GDP	RGDP	1968Q4	1981Q3	2009Q2	2009Q2
Unemployment rate	UNRATE	1968Q4	1981Q3	2009Q2	2009Q2
GDP price index	PGDP	1968Q4	1981Q3	NA	NA
CPI inflation	CPI	1981Q3	1981Q3	2005Q3	NA
T-bill rate	TBILL	1981Q3	1981Q3	2009Q2	2009Q2

Note: The table reports the first quarters in which SPF predictions become available in our data set for the stated variables and horizons. NA stands for not available. Prior to 1992, RGDP corresponds to real GNP, while PGDP corresponds to the GNP implicit deflator. The SPF solicits point forecasts for RGDP and PGDP in levels, which we convert to continuously compounded growth rates. The SPF also provides current year predictions that are, however, disregarded in our analysis due to overlap with the quarterly fixed-horizon predictions.

Table 2: Forecast accuracy (baseline MDS assumption)

	RMSE					CRPS				
Horizon h	RGDP	UNRATE	PGDP	CPI	TBILL	RGDP	UNRATE	PGDP	CPI	TBILL
0	1.00	1.00	1.00	1.00	1.00	1.00	1.09	1.02	1.04***	1.00
1	1.00	1.00	1.00	1.00	1.00	1.04*	1.06	1.02*	1.02*	1.10***
2	1.00	1.00	1.00	1.00	1.00	1.01	1.04	1.02*	1.01	1.04
3	1.00	1.00	1.00	1.00	1.00	1.02	1.01	1.03**	1.01	1.01
4	1.00	1.00	1.00	1.00	1.00	1.03*	0.99	1.01	1.02	0.99
5	1.00	1.01**	0.98	1.00	1.00	1.02	0.98	1.02	1.01	0.98
6	1.01*	1.02	1.04**	1.00	1.02	1.04**	0.98	1.09***	1.02	0.99
7	1.00	1.01*	1.04**	1.00	1.03	1.03*	0.98	1.12***	1.00	0.99
8	1.00	1.01	–	0.98	1.04	1.04*	0.99	–	0.99	0.99
9	1.00	1.01	–	1.00	1.05	1.04*	0.99	–	0.99	1.00
10	1.00	1.01	–	1.01	1.06	1.04**	1.00	–	0.99	1.01
11	1.00	1.01	–	1.00	1.06	1.05**	1.00	–	0.99	1.01
12	1.01	1.02	–	–	1.06	1.06***	1.00	–	–	1.00
13	1.01	1.01	–	–	1.05	1.06***	1.00	–	–	0.99
14	1.01	1.01	–	–	1.03	1.07***	0.99	–	–	0.97
15	1.01	1.01	–	–	1.02	1.06***	0.99	–	–	0.96

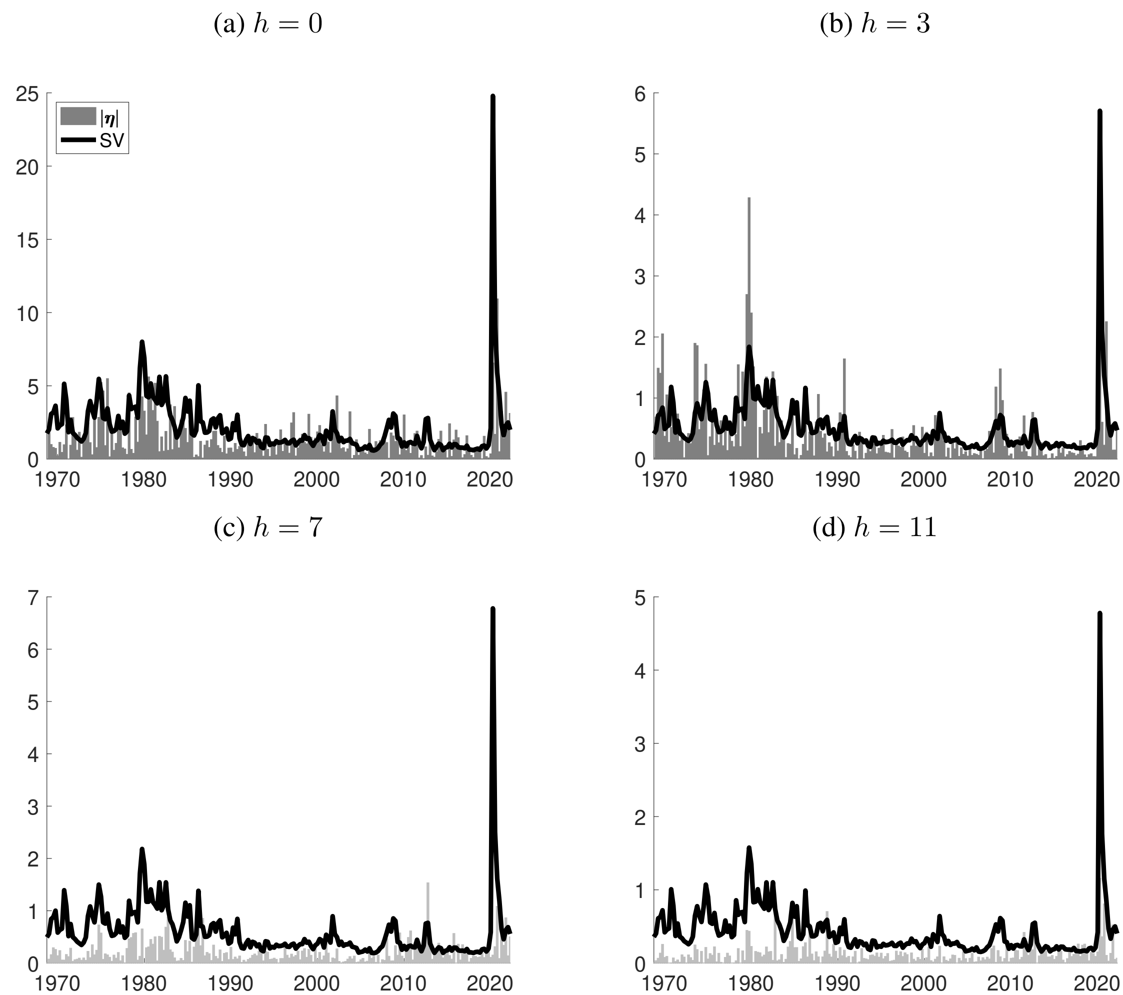
Note: Relative RMSE and CRPS of CONST model (with SV in denominator). Evaluation window from 1983Q3 through 2022Q2 (and as far as realized values are available). Variable mnemonics: RGDP denotes real growth, UNRATE the unemployment rate, PGDP changes in the GDP deflator, CPI inflation, and TBILL the interest rate on Treasury bills. (All growth rates are expressed as annualized percentage points of quarterly rates of change.) Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Table 3: Forecast accuracy without MDS assumption

	RMSE					CRPS				
Horizon h	RGDP	UNRATE	PGDP	CPI	TBILL	RGDP	UNRATE	PGDP	CPI	TBILL
0	1.00	1.01	0.99	0.93^{**}	0.92^{*}	1.01	1.01	1.00	0.95^{***}	0.94^{**}
1	1.02	1.00	1.00	1.00	1.00	1.02	1.01	0.99	1.00	1.00
2	1.00^{**}	1.01	0.99^{*}	1.00	1.00	1.00	1.01	0.99^{***}	1.00	0.99
3	1.00	1.01	0.99	1.00	0.98^{***}	1.01	1.01	0.99^{***}	1.00	0.98^{**}
4	1.00	1.00	0.99^{**}	1.00	0.98^{***}	1.01	1.00	0.98^{***}	1.00	0.99
5	1.00	1.01	0.96	1.00	0.97^{***}	1.01	1.01	0.96^{**}	1.00	0.98^{*}
6	1.00	1.01	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
7	1.00	1.01	1.03	1.00	1.00	1.00	1.00	1.02	0.99	1.00
8	1.00	1.00	–	0.99^{**}	1.00	1.00	1.00	–	0.98^{***}	1.00
9	1.00	1.00	–	0.99^{*}	1.00	1.00	1.00	–	0.98^{**}	1.01
10	1.00	1.00	–	1.00	1.01	1.00	1.01^{*}	–	0.99	1.01
11	1.00	1.00	–	0.99	1.01	1.00	1.01^{**}	–	0.98^{**}	1.01
12	1.00	1.00	–	–	1.01	1.00	1.01^{**}	–	–	1.02
13	1.00	1.00	–	–	1.02^{*}	1.00	1.01^{**}	–	–	1.03
14	1.00	1.00	–	–	1.02^{*}	0.99	1.01^{*}	–	–	1.03
15	1.00	1.00	–	–	1.03^{**}	0.99	1.01	–	–	1.03^{*}

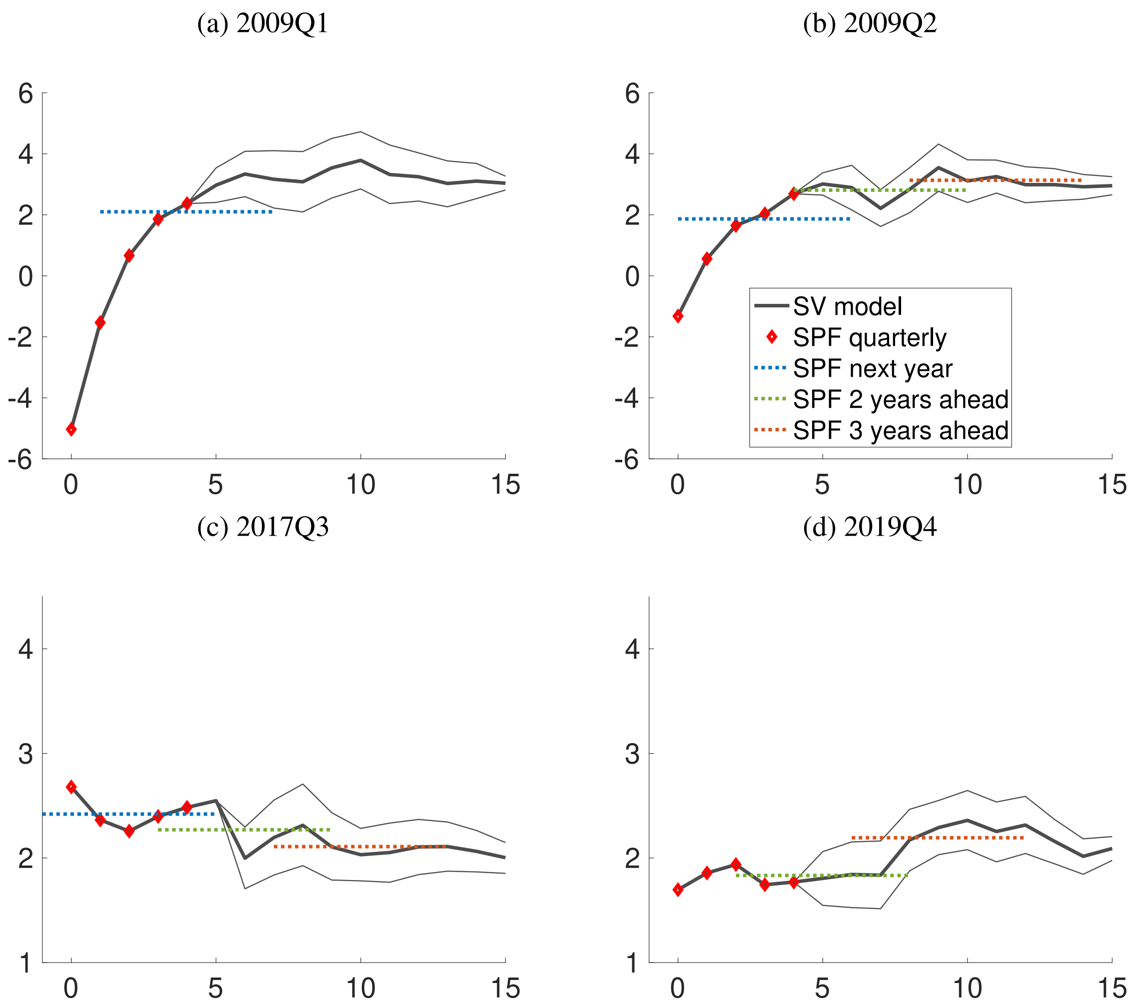
Note: Relative RMSE and CRPS of non-MDS (SV) model (with MDS (SV) in denominator). Evaluation window from 1983Q3 through 2022Q2 (and as far as realized values are available). Variable mnemonics: RGDP denotes real growth, UNRATE the unemployment rate, PGDP changes in the GDP deflator, CPI inflation, and TBILL the interest rate on Treasury bills. (All growth rates are expressed as annualized percentage points of quarterly rates of change.) Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Figure 1: Stochastic volatility estimates for GDP growth



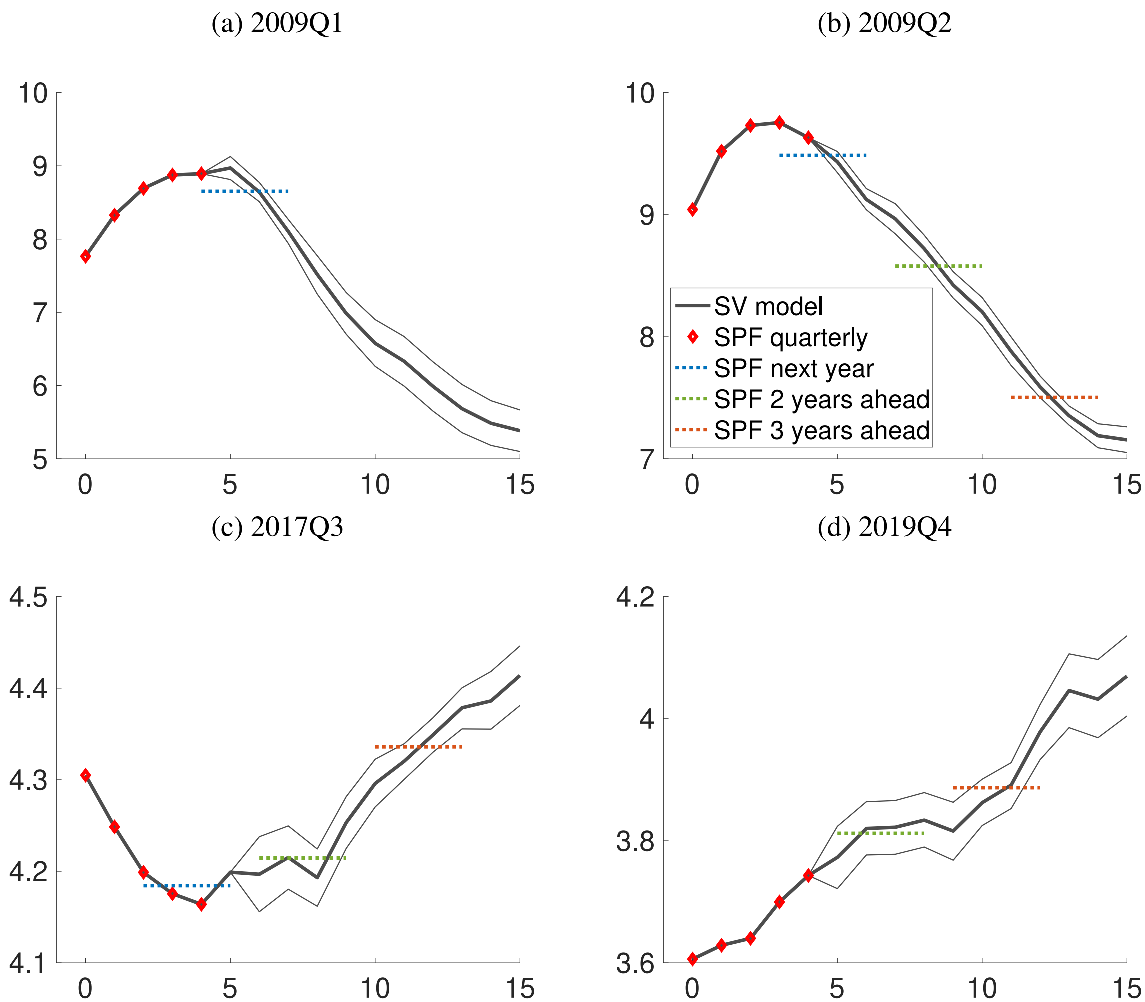
Notes: Stochastic volatility estimates for forecast updates, $\tilde{\boldsymbol{\eta}}_t$, of GDP growth at horizons $h = 0, 3, 7, 11$. Grey bars report the absolute values of the forecast errors at each horizon; darker grey bars for $h = 0$ and $h = 3$ indicate that forecast updates for those horizons are directly observed, whereas lighter grey bars for $h = 7$ and $h = 11$ reflect posterior medians obtained from our baseline SV model applied to the full data sample through 2022Q2. Black lines report (smoothed) volatility estimates obtained with the full sample of data (through 2022Q2).

Figure 2: Term structures of GDP growth expectations



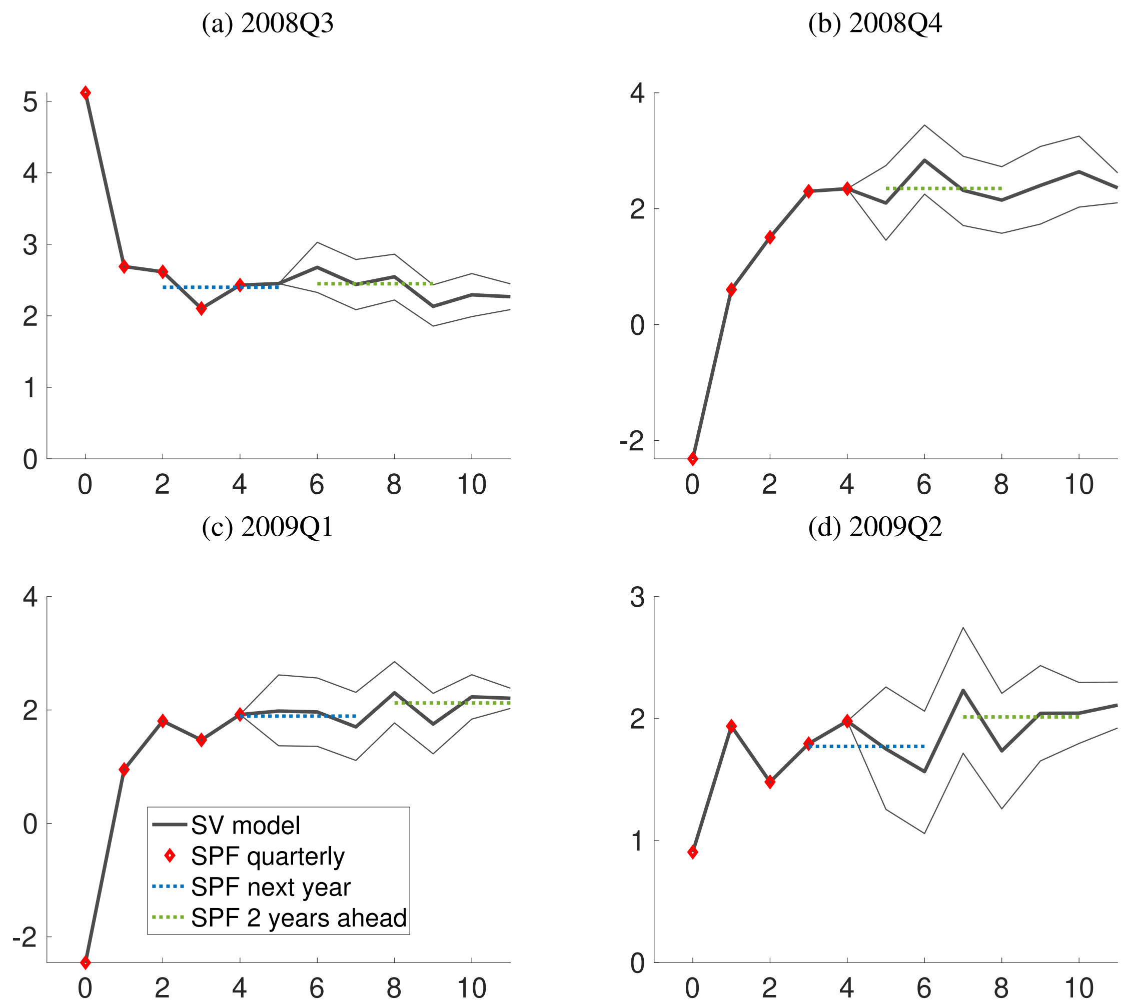
Notes: Term structures of GDP growth expectations, selected forecast origins. The black lines provide the real-time quarterly SPF-consistent forecast estimates, with 68 percent credible sets indicated with the thin lines. Red diamonds give the available quarterly fixed-horizon forecasts from the SPF, and dotted lines indicate the available annual fixed-event forecasts (the length of each line corresponds to the 7 quarters included in the tent-shaped linear mapping from annual-average to quarterly changes). Horizons (in quarters) are indicated on the horizontal axis.

Figure 3: Term structures of unemployment rate expectations



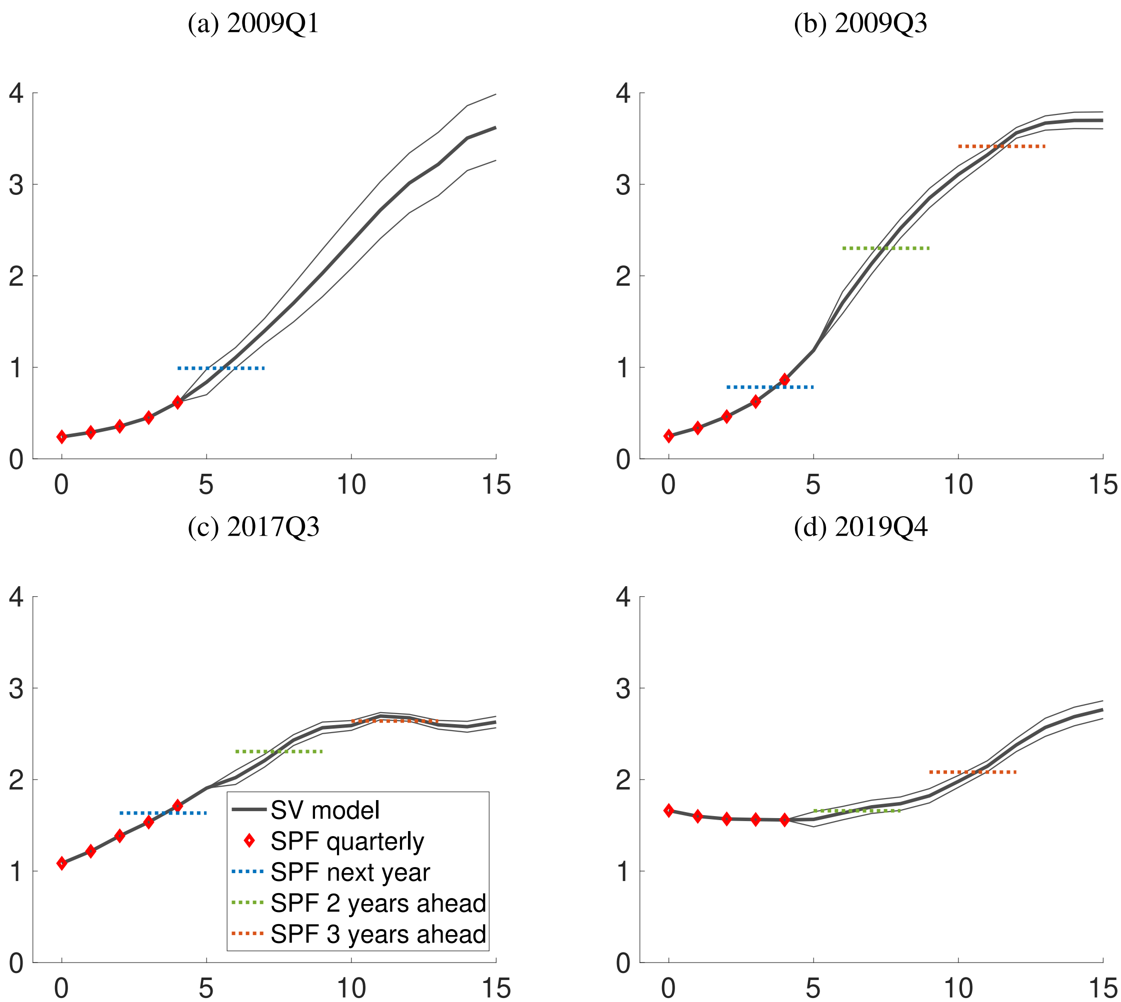
Notes: Term structures of unemployment rate expectations, selected forecast origins. The black lines provide the real-time quarterly SPF-consistent forecast estimates, with 68 percent credible sets indicated with the thin lines. Red diamonds give the available quarterly fixed-horizon forecasts from the SPF, and dotted lines indicate the available annual fixed-event forecasts (the length of each line corresponds to the 4 quarters included in the mapping from annual-average to quarterly levels). Horizons (in quarters) are indicated on the horizontal axis.

Figure 4: Term structures of CPI inflation expectations



Notes: Term structures of CPI inflation expectations, selected forecast origins. The black lines provide the real-time quarterly SPF-consistent forecast estimates, with 68 percent credible sets indicated with the thin lines. Red diamonds give the available quarterly fixed-horizon forecasts from the SPF, and dotted lines indicate the available annual fixed-event forecasts (the length of each line corresponds to the 4 quarters included in the mapping from annual-average to quarterly changes). Horizons (in quarters) are indicated on the horizontal axis.

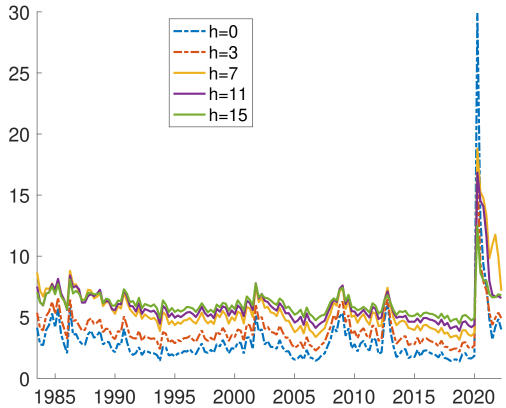
Figure 5: Term structures of T-bill rate expectations



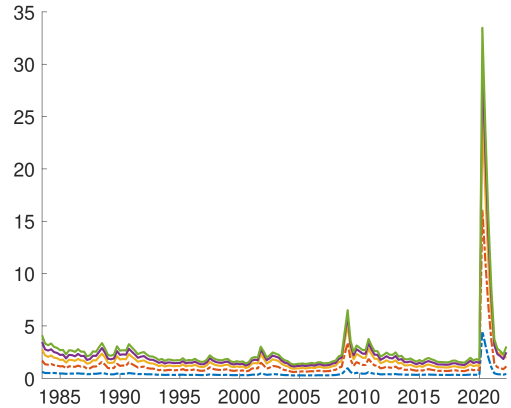
Notes: Term structures of T-bill rate expectations, selected forecast origins. The black lines provide the real-time quarterly SPF-consistent forecast estimates, with 68 percent credible sets indicated with the thin lines. Red diamonds give the available quarterly fixed-horizon forecasts from the SPF, and dotted lines indicate the available annual fixed-event forecasts (the length of each line corresponds to the 4 quarters included in the mapping from annual-average to quarterly levels). Horizons (in quarters) are indicated on the horizontal axis.

Figure 6: Term structures of uncertainty

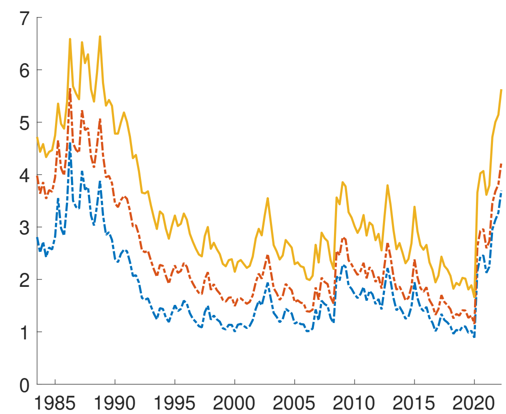
(a) GDP growth



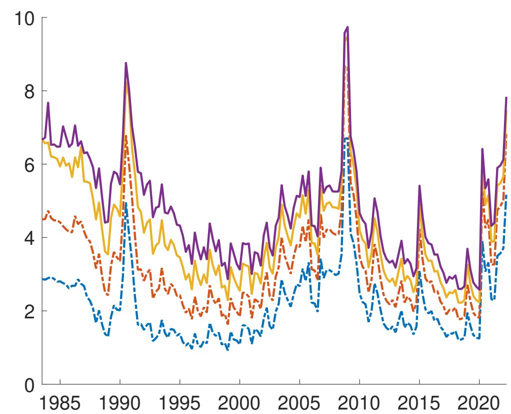
(b) Unemployment rate



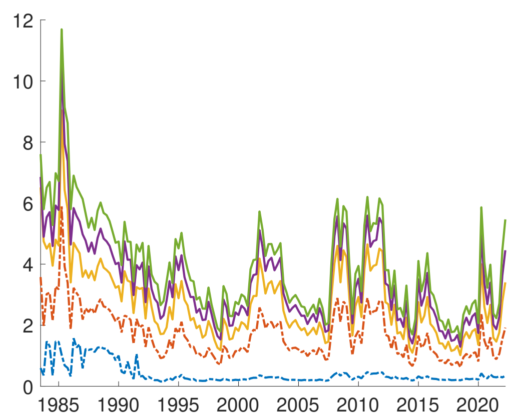
(c) GDP prices inflation



(d) CPI inflation

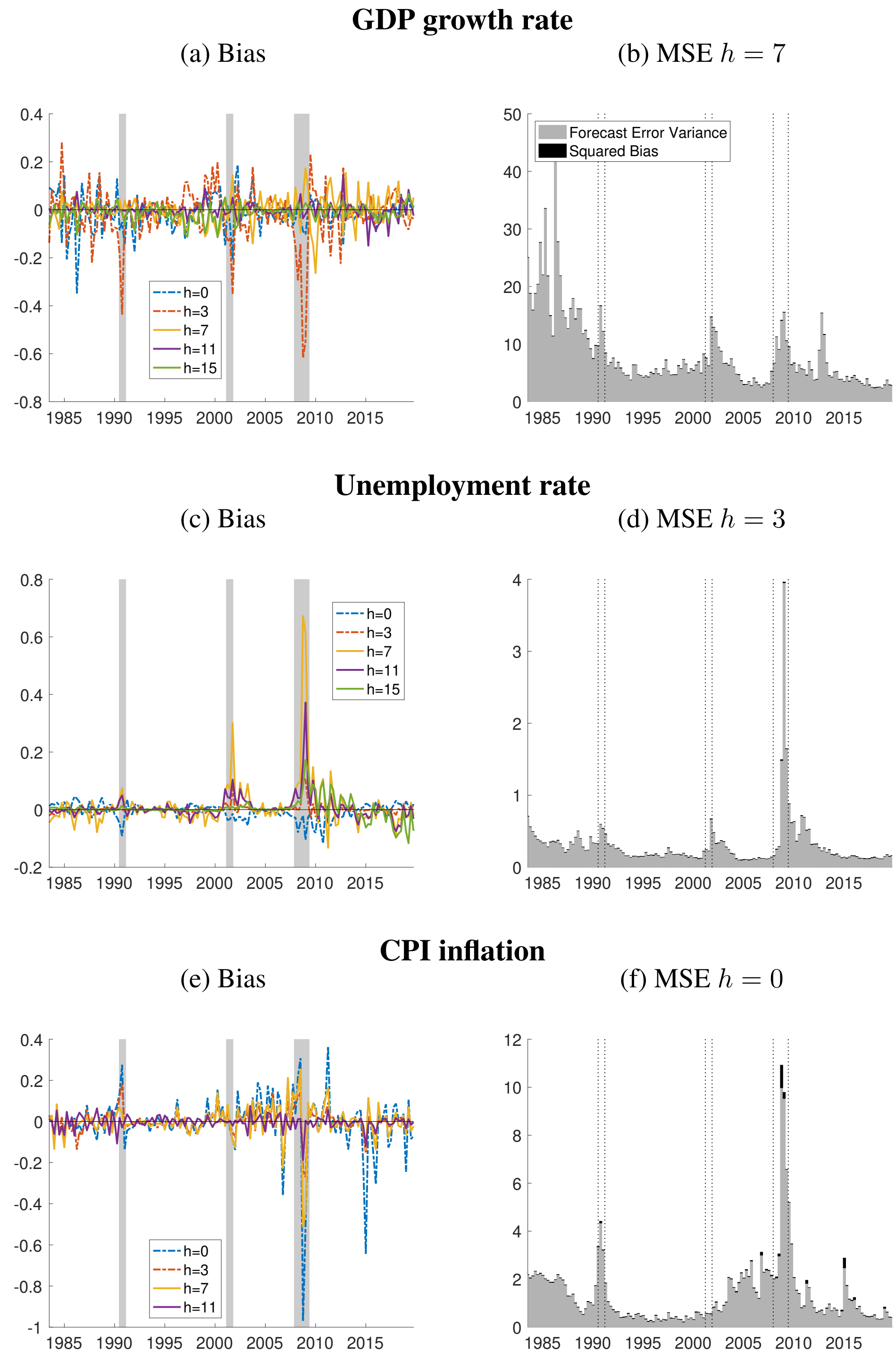


(e) T-bill rate



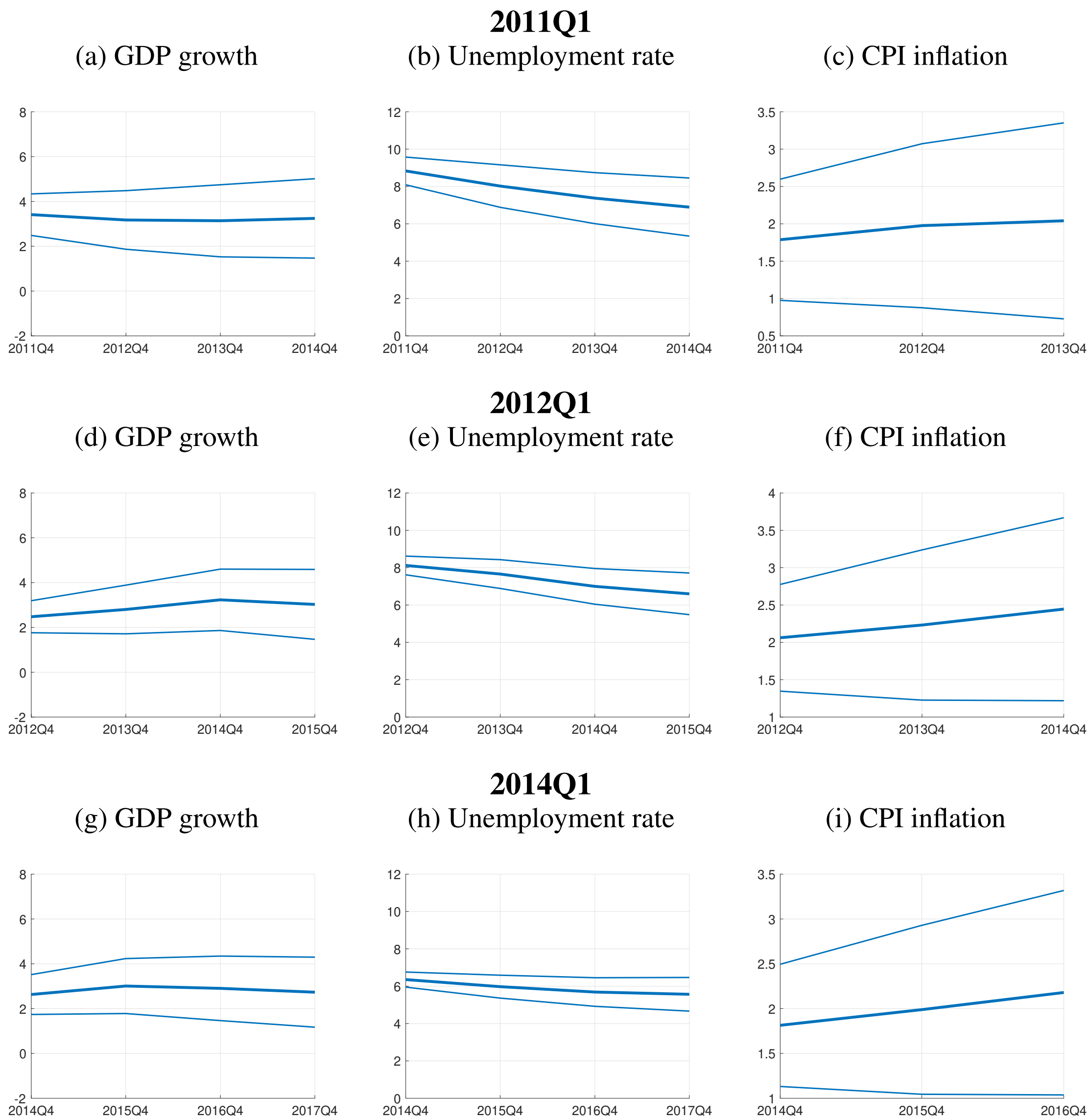
Notes: Term structures of uncertainty, measured by the width of 68% bands of predictive real-time densities from the baseline SV model.

Figure 7: Time-varying bias and MSE of SPF-consistent expectations in non-MDS model



Notes: Time-varying bias (left-column panels) and mean squared error (MSE) decompositions (right-column panels) from non-MDS model (with SV) for selected quarterly forecast horizons. The ex-ante MSE is decomposed into the sum of forecast error variance and squared bias. MSE decompositions are shown for those forecast horizons where contributions from squared bias are most visible. Results for alternative forecast horizons are shown in the supplementary online appendix. Shaded areas (left-column panels) and vertical dotted lines (right-column panels) demarcate NBER recessions.

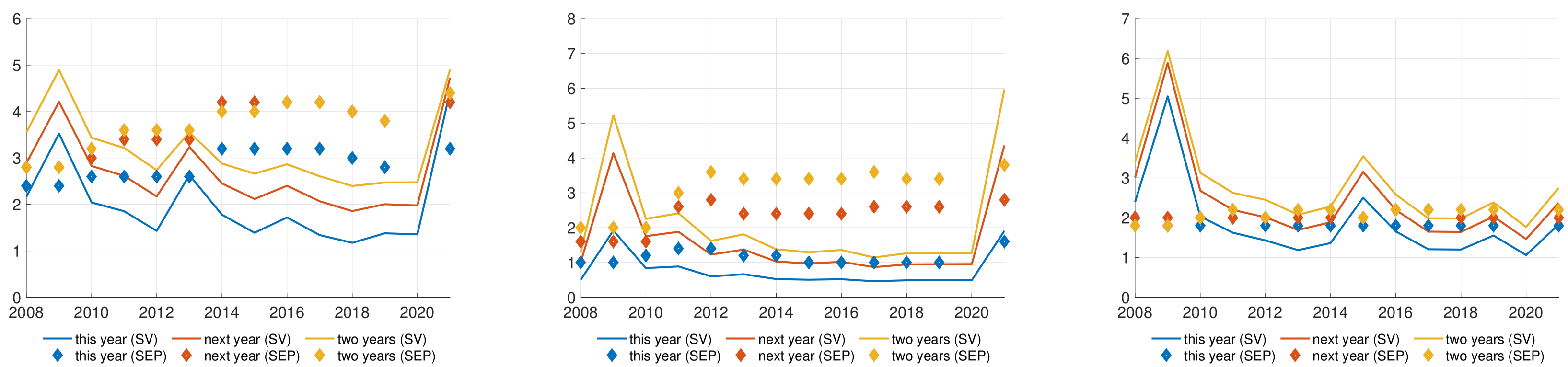
Figure 8: Fan charts for annual predictions from baseline SV model



Notes: Predictive means and 68% uncertainty bands generated by the SV model at various forecast origins. For GDP growth and CPI inflation, forecasts are for the average growth rate over the targeted and three previous quarters, whereas for the unemployment rate the forecast targets the quarterly level.

Figure 9: Uncertainty measures from SEP and baseline SV model, annual forecasts

(a) GDP growth (b) Unemployment rate (c) CPI inflation



Notes: Uncertainty measures as reported by the FOMC's SEP (diamonds) and generated from our SV model (solid lines). The SEP measure corresponds to two times the RMSE of historical errors of professional forecasters estimated as in Reifschneider and Tulip (2019), and reported in the SEP. The model-based measure is the width of the 68% uncertainty band at the given forecast horizon.